Auth0 Exercise - Data Engineer

Amol Surve

Exercise

Introduction

Attached is a reduced set of data, consisting on various tables:

dataset_emails: One row per email, represents an email of a person that signed up to Auth0 on https://auth0.com *dataset_pageviews*: One row per combination of url group and email, represents the # of visits of each email to a specific page group. *dataset_tenants*: One row per tenant, it's the account entity we use in Auth0 when you signup, other admins (emails) can be invited to the same tenant *dataset_relations*: One email can be associated with many tenants, and one tenant with many emails, so this is the relationship table that tells you which tenants are associated with which emails

This is real data from all emails signed up on July, but without any PII involved.

Requirements

You need to use R to solve this exercise, it's recommended to use R Studio to work on the solution.

Recommended packages: * dplyr - General data wrangling * tidyr - Data reshaping * ggplot2 - One of the best charting libraries, very flexible

To load the data, execute: load('exerciseData.rda')

Exercise

We want to analyze our customer data, and better understand them.

Please identify:

- The top 10 countries by active enterprise users
- The top 10 operating systems by open and total tickets
- The distribution of pageviews per group
- The distribution differences on pageviews per login method
- What differences do you find between developers and non-developers?
- What are the most and least used technologies in tenants?
- Please mark any inconsistencies you find in the data that you'd research further if you were working at Auth0
- Create at least 3 visualizations using R which you consider interesting for the provided data

```
# Loading the libraries
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

```
## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr

## Conflicts with tidy packages ----------------------------------------------

## filter(): dplyr, stats
## lag():    dplyr, stats
```

```
library(ggplot2)
library(reshape)
```

```
##
## Attaching package: 'reshape'

## The following objects are masked from 'package:tidyr':
##
##     expand, smiths

## The following object is masked from 'package:dplyr':
##
##     rename
```

```
library(data.table)
```

```
## ------------------------------------------------------------------------------

## data.table + dplyr code now lives in dtplyr.
## Please library(dtplyr)!

## ------------------------------------------------------------------------------

##
## Attaching package: 'data.table'

## The following object is masked from 'package:reshape':
##
##     melt

## The following object is masked from 'package:purrr':
##
##     transpose
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, last
```

```
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##     combine
```

```
#loading the data
load('exerciseData.rda')

#observing the dataset variables
str(dataset_emails)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    13127 obs. of  12 variables:
##  $ email_id                 : chr  "1" "2" "3" "4" ...
##  $ invited_to_dashboard     : logi  FALSE TRUE FALSE FALSE FALSE FALSE ...
##  $ country                  : chr  "Japan" "Australia" "Brazil" "Malaysia" ...
##  $ role                     : chr  "developer" "" "developer" "developer" ...
##  $ company_source           : chr  "None" "Salesforce" "None" "Salesforce" ...
##  $ company_employees        : int  NA 1058 NA 5 5 NA NA NA NA NA ...
##  $ source_group             : chr  "Direct" "Direct" "Direct" "Google" ...
##  $ sperating_system         : chr  "Windows 10" "Windows 10" "Android" "Mac OS X" ...
##  $ browser                  : chr  "Chrome" "Chrome" "Chrome Mobile" "Chrome" ...
##  $ dashboard_sessions       : int  2 4 9 4 2 7 1 1 7 2 ...
##  $ login_method             : chr  "auth0" "google-oauth2" "auth0" "auth0" ...
##  $ is_socialmedia_influencer: logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
```

```
str(dataset_tenants)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    10752 obs. of  15 variables:
##  $ tenant_id                          : chr  "1" "2" "3" "4" ...
##  $ region                             : chr  "europe" "master" "master" "australia" ...
##  $ paid                               : logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
##  $ active_users_last_30d              : int  NA NA NA NA 2 0 NA 0 1 1 ...
##  $ active_social_users_last_30d       : int  NA NA NA NA 0 0 NA 0 0 0 ...
##  $ active_enterprise_users_last_30d   : int  NA NA NA NA 0 0 NA 0 0 0 ...
##  $ total_apps                         : int  1 1 1 2 1 2 1 1 9 2 ...
##  $ connection_types                   : chr  "" "" "" "" ...
##  $ technologies_used                  : chr  "" "" "" "" ...
##  $ used_features                      : chr  "" "" "" "" ...
##  $ environment                        : chr  "" "" "" "" ...
##  $ account_open_tickets               : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ account_total_tickets              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ deployment_type                    : chr  "dm_public_cloud" "dm_public_cloud" "dm_public_cloud" "
##  $ anomaly_breached_password_detection: logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
```

```r
str(dataset_pageviews)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    153974 obs. of  3 variables:
##  $ email_id : chr  "4327" "7729" "1806" "1790" ...
##  $ type_page: chr  "Dashboard Clients" "Dashboard Clients" "Dashboard Clients" "Dashboard Rules" ...
##  $ visits   : int  516 303 365 226 80 1 6 64 8 11 ...
```

```r
str(dataset_relation)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    11757 obs. of  2 variables:
##  $ email_id : chr  "5393" "5501" "3232" "8619" ...
##  $ tenant_id: chr  "3128" "919" "3484" "4338" ...
```

```r
options(scipen=999)
```

Question 1

The top 10 countries by active enterprise users

```r
# Dataframe of country and top 10 active enterprise users
# Using dplyR functions like joins, select, mutate, group_by, summarise, filter(), etc. to subset the d
top_10_countries_active_enterprise_users <- dataset_emails %>%
  left_join(dataset_relation, by ="email_id") %>%
  left_join(dataset_tenants, by ="tenant_id") %>%
  select(country,active_enterprise_users_last_30d) %>%
  mutate(active_enterprise_users_last_30d =
  ifelse(is.na(active_enterprise_users_last_30d),0,active_enterprise_users_last_30d)) %>%
  group_by(country) %>%
  summarise(sum_of_active_enterprise_users = sum(active_enterprise_users_last_30d)) %>%
  arrange(desc(sum_of_active_enterprise_users)) %>%
  filter(country!="") %>% top_n(10)
```

```
## Selecting by sum_of_active_enterprise_users
```

```r
print(data.frame(top_10_countries_active_enterprise_users))
```

```
##           country sum_of_active_enterprise_users
## 1   United States                         410100
## 2  Czech Republic                          57866
## 3           India                          28956
## 4          France                          28933
## 5          Canada                          19604
## 6          Sweden                           2169
## 7       Australia                           1680
## 8  United Kingdom                           1560
## 9          Israel                           1003
## 10    New Zealand                            837
```

Question 2

The top 10 operating systems by open and total tickets

```
# Using dplyR functions like joins, select, mutate, group_by, summarise, filter(), etc. to subset the d
top_10_operating_systems_open_tickets <- dataset_emails %>%
  left_join(dataset_relation, by ="email_id") %>%
  left_join(dataset_tenants, by ="tenant_id") %>%
  select(sperating_system,account_open_tickets) %>%
  mutate(account_open_tickets = ifelse(is.na(account_open_tickets),0,account_open_tickets)) %>%
  group_by(sperating_system) %>%summarise(sum_of_account_open_tickets=sum(account_open_tickets)) %>%
  arrange(desc(sum_of_account_open_tickets)) %>%
  filter(sperating_system!="") %>% top_n(10)
```

```
## Selecting by sum_of_account_open_tickets
```

```
print(data.frame(top_10_operating_systems_open_tickets))
```

```
##    sperating_system sum_of_account_open_tickets
## 1          Mac OS X                         688
## 2        Windows 8.1                        248
## 3         Windows 10                        238
## 4          Windows 7                         89
## 5              Linux                         53
## 6             Ubuntu                         14
## 7            Android                          6
## 8                iOS                          5
## 9              Other                          5
## 10         Chrome OS                          2
## 11         Windows 8                          2
```

Question 3

The distribution of pageviews per group

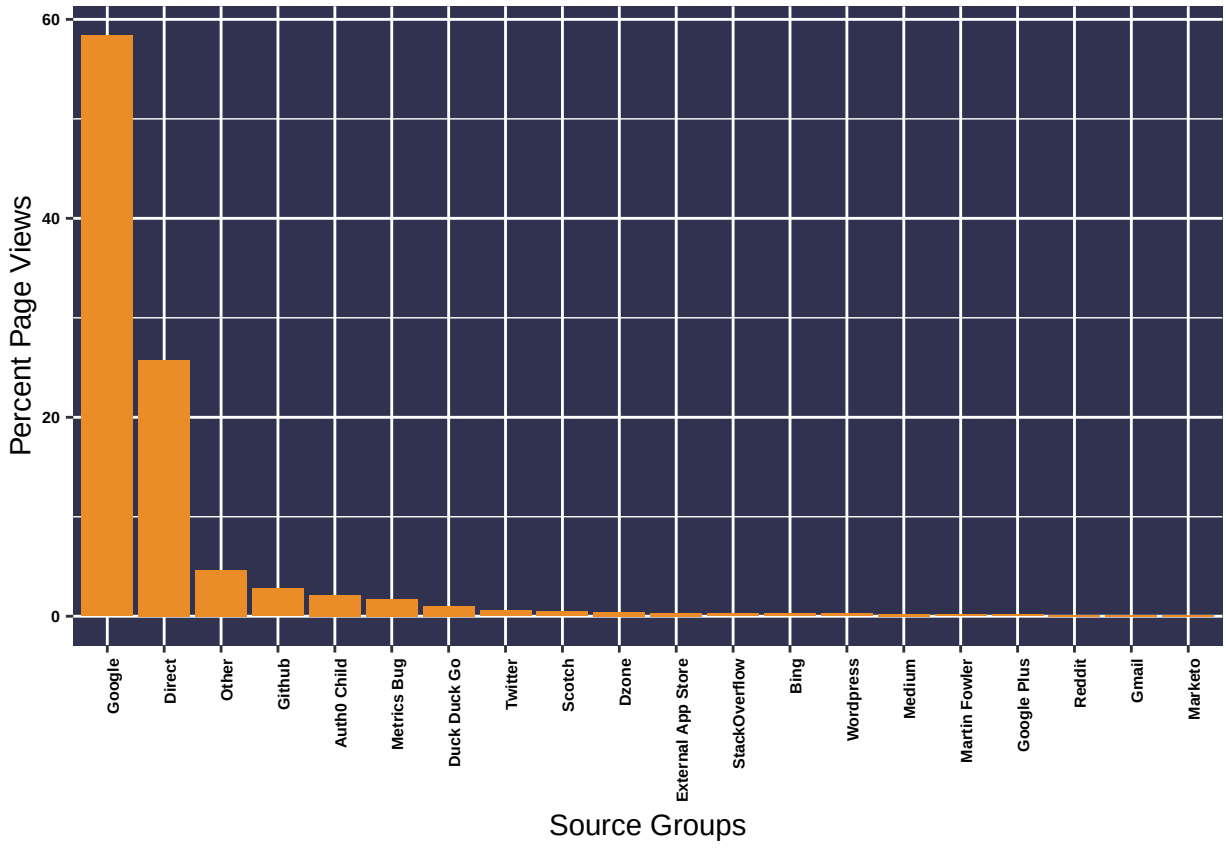
For this question, I have identified different factos related to the source groups and page type w.r.t. the page views. The following code shows the pageview distribution per page type for each group, as well as few other distributions like page views per group and page views per page type. I have visualized the results for few results to get the clearer understanding.

```
# Page Views group by source group and individual page type
page_views_per_group_with_page_type <- dataset_emails %>%
  left_join(dataset_pageviews, by ="email_id") %>%
  mutate(visits = ifelse(is.na(visits),0,visits))%>%
  group_by(source_group,type_page)

# Total Page Views grouped by source group
page_views_per_group <- dataset_emails %>%
  left_join(dataset_pageviews, by ="email_id") %>%
```

```
  mutate(visits = ifelse(is.na(visits),0,visits))%>%
  group_by(source_group) %>%
  summarise(views=sum(visits)) %>%
  arrange(desc(views))

# Percentage of page views per source group for each page type
page_views_per_group$views <- page_views_per_group$views  / sum(page_views_per_group$views)*100

# Plotting the top 20 results
visual_page_views_per_group <- dataset_emails %>%
  left_join(dataset_pageviews, by ="email_id") %>%
  mutate(visits = ifelse(is.na(visits),0,visits))%>%
  group_by(source_group) %>%
  summarise(views=sum(visits)) %>%
  arrange(desc(views)) %>%
  top_n(20)
```

```
## Selecting by views
```

```
visual_page_views_per_group$views <- visual_page_views_per_group$views  / sum(visual_page_views_per_grou

# Using ggplot
visual_page_views_per_group$source_group <- factor(visual_page_views_per_group$source_group, levels = vi
p1 <- ggplot(visual_page_views_per_group, aes(x = visual_page_views_per_group$source_group, y = visual_p
  geom_bar(stat = "identity",fill = "#ea8c27") +
  theme(panel.background = element_rect(fill = "#31324F",
                                colour = "#31324F",
                                size = 0.5, linetype = "solid"),  axis.text.x = element_text(face="bold
        axis.text.y = element_text(face="bold", color="black", size=6))+ xlab("Source Groups") +
  ylab("Percent Page Views")

# Displaying the plot of page views per source group
p1
```

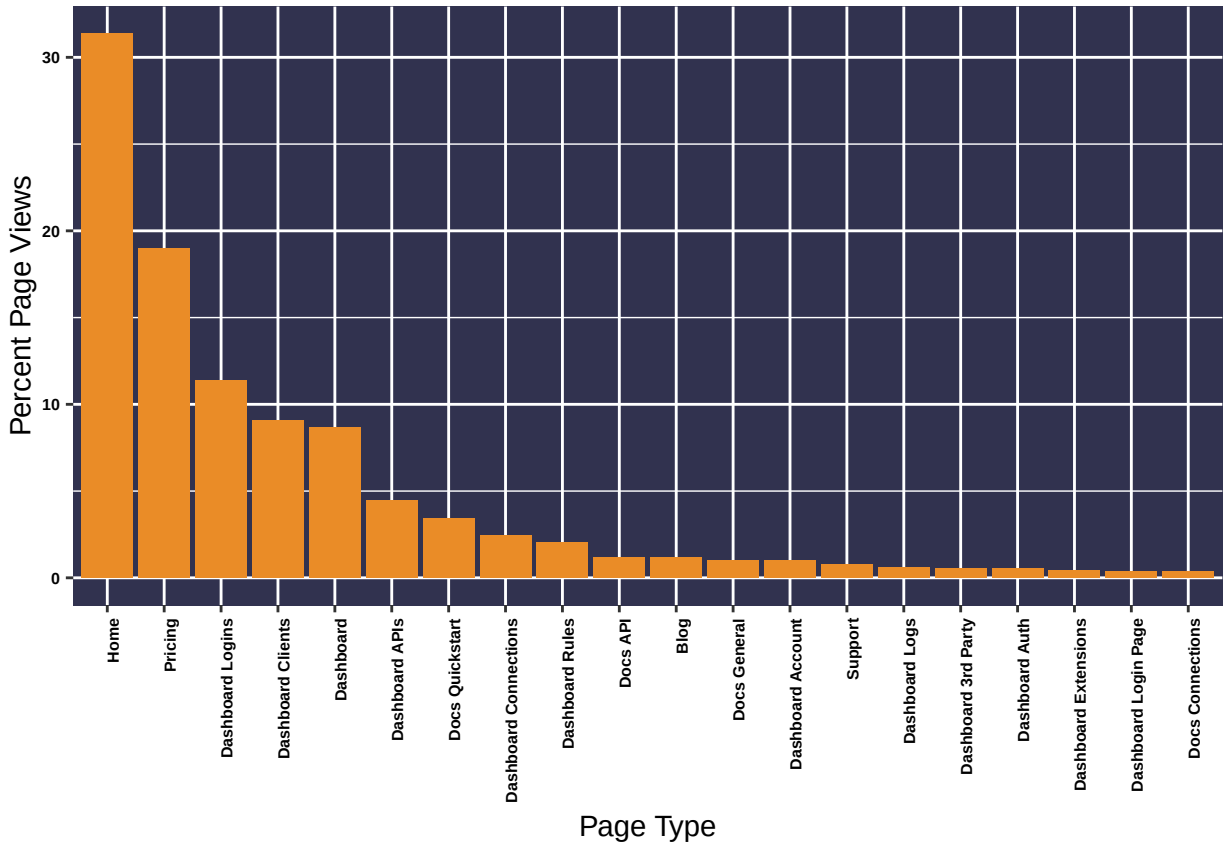



```
# Page views grouped by page types
page_views_per_page_type <- dataset_emails %>%
  left_join(dataset_pageviews, by ="email_id") %>%
  mutate(visits = ifelse(is.na(visits),0,visits))%>%
  group_by(type_page) %>%
  summarise(views=sum(visits)) %>%
  arrange(desc(views))

page_views_per_page_type$views <- page_views_per_page_type$views  / sum(page_views_per_page_type$views)

# Plotting the top 20 results
visual_page_views_per_page_type <- dataset_emails %>%
  left_join(dataset_pageviews, by ="email_id") %>%
  mutate(visits = ifelse(is.na(visits),0,visits))%>%
  group_by(type_page) %>%
  summarise(views=sum(visits)) %>%
  arrange(desc(views)) %>%
  top_n(20)
```

```
## Selecting by views
```

```
visual_page_views_per_page_type$views <- visual_page_views_per_page_type$views  / sum(visual_page_views

# Using ggplot
visual_page_views_per_page_type$type_page <- factor(visual_page_views_per_page_type$type_page, levels =
p1 <- ggplot(visual_page_views_per_page_type, aes(x = visual_page_views_per_page_type$type_page, y = vis
```

```
  geom_bar(stat = "identity",fill = "#ea8c27") +
  theme(panel.background = element_rect(fill = "#31324F",
                                        colour = "#31324F",
                                        size = 0.5, linetype = "solid"),  axis.text.x = element_text(face="bold
        axis.text.y = element_text(face="bold", color="black", size=6))+ xlab("Page Type") +
  ylab("Percent Page Views")

# displaying the plot of page views per page type without considering the group
p1
```



Question 4

The distribution differences on pageviews per login method

For this question, the distribution for the page views per page type w.r.t every login method is shown. Then, further the distribution of page views grouped w.r.t login_method is visualized without considerig the granularities of the page types for each method.

```
# Identifying the distribution of page view per login method for each page type
page_per_views_login_method <- dataset_emails %>%
  left_join(dataset_pageviews, by ="email_id") %>%
  mutate(visits = ifelse(is.na(visits),0,visits))%>%
  select(login_method, type_page, visits) %>%
  group_by(login_method, type_page) %>%
```

```
  summarize(total_visits = sum(visits))

print(data.frame(page_per_views_login_method))
```

```
##             login_method                  type_page total_visits
## 1                   adfs                       Blog           1
## 2                   adfs                  Dashboard         284
## 3                   adfs          Dashboard Account           9
## 4                   adfs             Dashboard Auth          29
## 5                   adfs          Dashboard Clients         247
## 6                   adfs       Dashboard Connections         114
## 7                   adfs            Dashboard Emails           6
## 8                   adfs            Dashboard invite           5
## 9                   adfs            Dashboard Logins          80
## 10                  adfs             Dashboard Rules          66
## 11                  adfs                 Docs Emails           1
## 12                  adfs             Docs Extensions           1
## 13                  adfs                Docs General           9
## 14                  adfs                    Docs IdP           1
## 15                  adfs             Docs Protocols           1
## 16                  adfs             Docs Quickstart          19
## 17                  adfs                       Home         687
## 18                  adfs                    Support          16
## 19                 auth0                      About         366
## 20                 auth0                 AuthCatalog        4457
## 21                 auth0                Availability         135
## 22                 auth0                       Blog       23449
## 23                 auth0                  Changelog          32
## 24                 auth0                 Compliance          83
## 25                 auth0                  Dashboard      200719
## 26                 auth0         Dashboard 3rd Party       12684
## 27                 auth0          Dashboard Account       24633
## 28                 auth0             Dashboard APIs       99695
## 29                 auth0             Dashboard Auth       12688
## 30                 auth0 Dashboard Breached Password        2486
## 31                 auth0          Dashboard Clients      204611
## 32                 auth0       Dashboard Connections       57428
## 33                 auth0            Dashboard Emails        9129
## 34                 auth0        Dashboard Extensions        9929
## 35                 auth0            Dashboard invite        5068
## 36                 auth0        Dashboard Login Page        9111
## 37                 auth0            Dashboard Logins      281398
## 38                 auth0              Dashboard Logs       15600
## 39                 auth0               Dashboard MFA        5838
## 40                 auth0             Dashboard Rules       48832
## 41                 auth0                 Docs Addons         186
## 42                 auth0                    Docs API       26834
## 43                 auth0              Docs Appliance         736
## 44                 auth0                    Docs CMS         326
## 45                 auth0          Docs Connections        7686
## 46                 auth0            Docs Connector         406
## 47                 auth0                 Docs Emails        1138
## 48                 auth0             Docs Extensions        1909
```

```
## 49                auth0                Docs General        22579
## 50                auth0           Docs hosted pages         1171
## 51                auth0                   Docs i18n           40
## 52                auth0                    Docs IdP         4046
## 53                auth0           Docs Integrations         1047
## 54                auth0         Docs Library Auth0js         2609
## 55                auth0            Docs Library Lock         8275
## 56                auth0   Docs Library Lock Android          330
## 57                auth0       Docs Library Lock iOS          522
## 58                auth0                    Docs MFA         1016
## 59                auth0              Docs Protocols         2742
## 60                auth0             Docs Quickstart        73299
## 61                auth0                  Docs Rules         2010
## 62                auth0              Docs Scenarios         1705
## 63                auth0               Docs Services          120
## 64                auth0                Docs Support          320
## 65                auth0              Docs Tutorials         1966
## 66                auth0                     E-Books          956
## 67                auth0                      Events            1
## 68                auth0                   Glosaries           30
## 69                auth0                      Hiring          378
## 70                auth0                        Home       669770
## 71                auth0                Japanese Site          438
## 72                auth0           Learn Case Study           68
## 73                auth0           Learn Competition          123
## 74                auth0                   Learn CSM           13
## 75                auth0                 Learn Guide          712
## 76                auth0                 Learn Offer           26
## 77                auth0              Learn Tutorial         1373
## 78                auth0              Learn Use Case          355
## 79                auth0             Learn Verticals           16
## 80                auth0                  OpenSource          199
## 81                auth0                       Other          208
## 82                auth0                    Partners           10
## 83                auth0                       Press           57
## 84                auth0                     Pricing       445419
## 85                auth0                    Products         1736
## 86                auth0                   Resources          258
## 87                auth0                       Rules          378
## 88                auth0                    Security          306
## 89                auth0                  Signup page         5030
## 90                auth0                    Solutions          644
## 91                auth0                      Support        19616
## 92                auth0                   University          414
## 93                auth0                 User Program           71
## 94                auth0           Website Education         1145
## 95                auth0                         <NA>            0
## 96               github                        About           88
## 97               github                   AuthCatalog         1654
## 98               github                  Availability           22
## 99               github                          Blog        10012
## 100              github                     Changelog            4
## 101              github                    Compliance            9
## 102              github                     Dashboard        47211
```

```
## 103              github         Dashboard 3rd Party         3365
## 104              github           Dashboard Account         5221
## 105              github              Dashboard APIs        25127
## 106              github              Dashboard Auth         2047
## 107              github Dashboard Breached Password          567
## 108              github           Dashboard Clients        54241
## 109              github       Dashboard Connections        15359
## 110              github            Dashboard Emails         1029
## 111              github        Dashboard Extensions         2167
## 112              github            Dashboard invite          692
## 113              github        Dashboard Login Page         2043
## 114              github            Dashboard Logins        41723
## 115              github              Dashboard Logs         2639
## 116              github               Dashboard MFA         1352
## 117              github             Dashboard Rules         9952
## 118              github                 Docs Addons           76
## 119              github                    Docs API         6795
## 120              github              Docs Appliance          240
## 121              github                    Docs CMS           15
## 122              github            Docs Connections         2067
## 123              github              Docs Connector           56
## 124              github                 Docs Emails          220
## 125              github             Docs Extensions          364
## 126              github                Docs General         6360
## 127              github           Docs hosted pages          347
## 128              github                   Docs i18n            8
## 129              github                    Docs IdP          966
## 130              github           Docs Integrations          485
## 131              github        Docs Library Auth0js          796
## 132              github           Docs Library Lock         2064
## 133              github   Docs Library Lock Android           98
## 134              github       Docs Library Lock iOS          108
## 135              github                   Docs MFA          155
## 136              github              Docs Protocols          440
## 137              github             Docs Quickstart        21695
## 138              github                  Docs Rules          425
## 139              github             Docs Scenarios          448
## 140              github              Docs Services           16
## 141              github              Docs Support            40
## 142              github             Docs Tutorials          542
## 143              github                    E-Books         1103
## 144              github                     Events            1
## 145              github                  Glosaries           17
## 146              github                     Hiring          238
## 147              github                       Home       205418
## 148              github               Japanese Site          330
## 149              github            Learn Case Study            4
## 150              github           Learn Competition           45
## 151              github                   Learn CSM            4
## 152              github                Learn Guide          153
## 153              github                Learn Offer           13
## 154              github              Learn Tutorial          501
## 155              github              Learn Use Case           72
## 156              github             Learn Verticals            2
```

```
## 157              github                  OpenSource          66
## 158              github                       Other          55
## 159              github                       Press           5
## 160              github                     Pricing      109736
## 161              github                    Products         488
## 162              github                   Resources          48
## 163              github                       Rules          77
## 164              github                    Security          89
## 165              github                  Signup page        1593
## 166              github                    Solutions         111
## 167              github                     Support        3156
## 168              github                  University          39
## 169              github                User Program          29
## 170              github           Website Education         384
## 171              github                       <NA>           0
## 172        google-apps                       About          57
## 173        google-apps                   AuthCatalog          30
## 174        google-apps                  Availability          10
## 175        google-apps                       Blog         171
## 176        google-apps                  Changelog           3
## 177        google-apps                 Compliance           8
## 178        google-apps                  Dashboard        1035
## 179        google-apps          Dashboard 3rd Party          46
## 180        google-apps            Dashboard Account         207
## 181        google-apps               Dashboard APIs         607
## 182        google-apps               Dashboard Auth          40
## 183        google-apps Dashboard Breached Password          16
## 184        google-apps            Dashboard Clients        1001
## 185        google-apps        Dashboard Connections         601
## 186        google-apps             Dashboard Emails          60
## 187        google-apps         Dashboard Extensions         108
## 188        google-apps             Dashboard invite          38
## 189        google-apps         Dashboard Login Page         102
## 190        google-apps             Dashboard Logins         985
## 191        google-apps               Dashboard Logs         106
## 192        google-apps                Dashboard MFA          96
## 193        google-apps              Dashboard Rules         451
## 194        google-apps                  Docs Addons           1
## 195        google-apps                     Docs API         456
## 196        google-apps               Docs Appliance          35
## 197        google-apps             Docs Connections         156
## 198        google-apps               Docs Connector           9
## 199        google-apps                  Docs Emails           1
## 200        google-apps              Docs Extensions          29
## 201        google-apps                 Docs General         437
## 202        google-apps            Docs hosted pages          12
## 203        google-apps                      Docs IdP          44
## 204        google-apps            Docs Integrations           3
## 205        google-apps         Docs Library Auth0js          21
## 206        google-apps            Docs Library Lock          58
## 207        google-apps   Docs Library Lock Android          14
## 208        google-apps        Docs Library Lock iOS          13
## 209        google-apps                     Docs MFA          44
## 210        google-apps               Docs Protocols          51
```

```
## 211          google-apps             Docs Quickstart        2187
## 212          google-apps                  Docs Rules          28
## 213          google-apps              Docs Scenarios          21
## 214          google-apps               Docs Services          17
## 215          google-apps                Docs Support          62
## 216          google-apps              Docs Tutorials          50
## 217          google-apps                   Glosaries           2
## 218          google-apps                      Hiring         112
## 219          google-apps                        Home       10093
## 220          google-apps           Learn Case Study           4
## 221          google-apps                  Learn CSM           1
## 222          google-apps                Learn Guide          11
## 223          google-apps                Learn Offer           6
## 224          google-apps              Learn Tutorial           7
## 225          google-apps              Learn Use Case          11
## 226          google-apps                  OpenSource           3
## 227          google-apps                       Other           7
## 228          google-apps                    Partners           1
## 229          google-apps                       Press           4
## 230          google-apps                     Pricing        9706
## 231          google-apps                    Products          26
## 232          google-apps                   Resources           4
## 233          google-apps                       Rules           2
## 234          google-apps                    Security          17
## 235          google-apps                  Signup page           6
## 236          google-apps                   Solutions           2
## 237          google-apps                     Support         440
## 238          google-apps                  University          23
## 239          google-apps                User Program           1
## 240          google-apps           Website Education           2
## 241          google-apps                       <NA>           0
## 242       google-oauth2                      About         220
## 243       google-oauth2                  AuthCatalog        5299
## 244       google-oauth2                 Availability         110
## 245       google-oauth2                         Blog       20683
## 246       google-oauth2                    Changelog           5
## 247       google-oauth2                   Compliance          32
## 248       google-oauth2                    Dashboard      138030
## 249       google-oauth2           Dashboard 3rd Party        8838
## 250       google-oauth2             Dashboard Account       15323
## 251       google-oauth2                Dashboard APIs       73195
## 252       google-oauth2                Dashboard Auth        8423
## 253       google-oauth2 Dashboard Breached Password        1533
## 254       google-oauth2              Dashboard Clients      146892
## 255       google-oauth2         Dashboard Connections       36838
## 256       google-oauth2              Dashboard Emails        4607
## 257       google-oauth2          Dashboard Extensions        6500
## 258       google-oauth2              Dashboard invite        2479
## 259       google-oauth2          Dashboard Login Page        6077
## 260       google-oauth2              Dashboard Logins      186491
## 261       google-oauth2                Dashboard Logs        7511
## 262       google-oauth2                 Dashboard MFA        4450
## 263       google-oauth2               Dashboard Rules       32607
## 264       google-oauth2                   Docs Addons         200
```

```
## 265        google-oauth2                  Docs API        20060
## 266        google-oauth2             Docs Appliance          456
## 267        google-oauth2                  Docs CMS          181
## 268        google-oauth2           Docs Connections         6805
## 269        google-oauth2             Docs Connector          238
## 270        google-oauth2                Docs Emails          819
## 271        google-oauth2            Docs Extensions         1448
## 272        google-oauth2               Docs General        16469
## 273        google-oauth2          Docs hosted pages          965
## 274        google-oauth2                  Docs i18n           19
## 275        google-oauth2                  Docs IdP         3009
## 276        google-oauth2          Docs Integrations          953
## 277        google-oauth2       Docs Library Auth0js         1905
## 278        google-oauth2          Docs Library Lock         5804
## 279        google-oauth2   Docs Library Lock Android          542
## 280        google-oauth2       Docs Library Lock iOS          483
## 281        google-oauth2                  Docs MFA          771
## 282        google-oauth2             Docs Protocols         1527
## 283        google-oauth2            Docs Quickstart        55666
## 284        google-oauth2                 Docs Rules         1563
## 285        google-oauth2             Docs Scenarios         1378
## 286        google-oauth2              Docs Services           73
## 287        google-oauth2              Docs Support          146
## 288        google-oauth2             Docs Tutorials         1377
## 289        google-oauth2                    E-Books         1977
## 290        google-oauth2                  Glosaries           15
## 291        google-oauth2                     Hiring          292
## 292        google-oauth2                       Home       535892
## 293        google-oauth2               Japanese Site          270
## 294        google-oauth2            Learn Case Study           35
## 295        google-oauth2           Learn Competition           86
## 296        google-oauth2                    Learn CSM           15
## 297        google-oauth2                 Learn Guide          483
## 298        google-oauth2                 Learn Offer            1
## 299        google-oauth2              Learn Tutorial         1348
## 300        google-oauth2              Learn Use Case          212
## 301        google-oauth2             Learn Verticals           10
## 302        google-oauth2                  OpenSource          143
## 303        google-oauth2                       Other          176
## 304        google-oauth2                       Press           13
## 305        google-oauth2                     Pricing       294603
## 306        google-oauth2                    Products         1777
## 307        google-oauth2                   Resources          135
## 308        google-oauth2                       Rules          239
## 309        google-oauth2                    Security          247
## 310        google-oauth2                  Signup page         4029
## 311        google-oauth2                   Solutions          465
## 312        google-oauth2                     Support        10435
## 313        google-oauth2                  University          227
## 314        google-oauth2                User Program           33
## 315        google-oauth2           Website Education          848
## 316        google-oauth2                        <NA>            0
## 317 hus_17ipdhxFxq0bZHTK                        <NA>            0
## 318 hus_1W92YUlTcmdNJRbJ                        <NA>            0
```

```
## 319 hus_2dor8KMpxRrz1P2U                      Blog           2
## 320 hus_2dor8KMpxRrz1P2U                 Dashboard          34
## 321 hus_2dor8KMpxRrz1P2U       Dashboard 3rd Party           2
## 322 hus_2dor8KMpxRrz1P2U           Dashboard APIs          12
## 323 hus_2dor8KMpxRrz1P2U Dashboard Breached Password           5
## 324 hus_2dor8KMpxRrz1P2U         Dashboard Clients          45
## 325 hus_2dor8KMpxRrz1P2U          Dashboard Logins          30
## 326 hus_2dor8KMpxRrz1P2U            Dashboard Logs           1
## 327 hus_2dor8KMpxRrz1P2U                 Docs API           4
## 328 hus_2dor8KMpxRrz1P2U          Docs Connections           1
## 329 hus_2dor8KMpxRrz1P2U              Docs General           5
## 330 hus_2dor8KMpxRrz1P2U                  Docs IdP           2
## 331 hus_2dor8KMpxRrz1P2U         Docs Integrations           1
## 332 hus_2dor8KMpxRrz1P2U           Docs Quickstart          10
## 333 hus_2dor8KMpxRrz1P2U                      Home         112
## 334 hus_2dor8KMpxRrz1P2U           Learn Tutorial           1
## 335 hus_2dor8KMpxRrz1P2U                  Pricing           1
## 336 hus_2dor8KMpxRrz1P2U                Solutions           1
## 337 hus_3u5RLFbrKocGJ03E                     <NA>           0
## 338 hus_4GidGRJeWhlO9rYq                     <NA>           0
## 339 hus_4HH22MMDErpFuzOg                     <NA>           0
## 340 hus_8RtiPmhaUhhnde5o                     <NA>           0
## 341 hus_9OoWnVnSRKNAazYk                     <NA>           0
## 342 hus_A1ZjEdFrJUeCAOUt                     <NA>           0
## 343 hus_aETmN8sUX5uzHjRZ             Availability           1
## 344 hus_aETmN8sUX5uzHjRZ                      Blog          11
## 345 hus_aETmN8sUX5uzHjRZ                 Dashboard         331
## 346 hus_aETmN8sUX5uzHjRZ       Dashboard 3rd Party          56
## 347 hus_aETmN8sUX5uzHjRZ         Dashboard Account          18
## 348 hus_aETmN8sUX5uzHjRZ           Dashboard APIs         190
## 349 hus_aETmN8sUX5uzHjRZ           Dashboard Auth           4
## 350 hus_aETmN8sUX5uzHjRZ Dashboard Breached Password           5
## 351 hus_aETmN8sUX5uzHjRZ         Dashboard Clients         424
## 352 hus_aETmN8sUX5uzHjRZ      Dashboard Connections         173
## 353 hus_aETmN8sUX5uzHjRZ          Dashboard Emails          16
## 354 hus_aETmN8sUX5uzHjRZ       Dashboard Extensions          90
## 355 hus_aETmN8sUX5uzHjRZ          Dashboard invite           5
## 356 hus_aETmN8sUX5uzHjRZ       Dashboard Login Page          12
## 357 hus_aETmN8sUX5uzHjRZ          Dashboard Logins         968
## 358 hus_aETmN8sUX5uzHjRZ            Dashboard Logs          21
## 359 hus_aETmN8sUX5uzHjRZ             Dashboard MFA          12
## 360 hus_aETmN8sUX5uzHjRZ           Dashboard Rules         246
## 361 hus_aETmN8sUX5uzHjRZ                 Docs API          43
## 362 hus_aETmN8sUX5uzHjRZ          Docs Connections          33
## 363 hus_aETmN8sUX5uzHjRZ            Docs Connector           1
## 364 hus_aETmN8sUX5uzHjRZ               Docs Emails           3
## 365 hus_aETmN8sUX5uzHjRZ           Docs Extensions          23
## 366 hus_aETmN8sUX5uzHjRZ              Docs General          58
## 367 hus_aETmN8sUX5uzHjRZ        Docs hosted pages           3
## 368 hus_aETmN8sUX5uzHjRZ                  Docs IdP          21
## 369 hus_aETmN8sUX5uzHjRZ       Docs Library Auth0js          15
## 370 hus_aETmN8sUX5uzHjRZ         Docs Library Lock           8
## 371 hus_aETmN8sUX5uzHjRZ            Docs Protocols           9
## 372 hus_aETmN8sUX5uzHjRZ           Docs Quickstart          66
```

```
## 373 hus_aETmN8sUX5uzHjRZ                 Docs Rules          13
## 374 hus_aETmN8sUX5uzHjRZ              Docs Scenarios           9
## 375 hus_aETmN8sUX5uzHjRZ              Docs Tutorials          15
## 376 hus_aETmN8sUX5uzHjRZ                        Home         351
## 377 hus_aETmN8sUX5uzHjRZ              Learn Tutorial           1
## 378 hus_aETmN8sUX5uzHjRZ                     Pricing           1
## 379 hus_aETmN8sUX5uzHjRZ                       Rules           1
## 380 hus_aETmN8sUX5uzHjRZ                 Signup page           1
## 381 hus_aETmN8sUX5uzHjRZ                     Support          13
## 382 hus_AL0VPy3S3kjfPVpe                        <NA>           0
## 383 hus_Am0LAlPGoDWLb64i                        <NA>           0
## 384 hus_AZHbtPWE0ZWphL41                        <NA>           0
## 385 hus_b9OSugfs0nkQnCoU                        <NA>           0
## 386 hus_BOw0mvqcPGZMxLpm                        <NA>           0
## 387 hus_BTgEh70v0nbWBPlR                        <NA>           0
## 388 hus_DTiKzwvw1Feny8xs                 AuthCatalog           4
## 389 hus_DTiKzwvw1Feny8xs                        Blog          16
## 390 hus_DTiKzwvw1Feny8xs                   Dashboard          64
## 391 hus_DTiKzwvw1Feny8xs           Dashboard 3rd Party          90
## 392 hus_DTiKzwvw1Feny8xs             Dashboard Account           3
## 393 hus_DTiKzwvw1Feny8xs                Dashboard APIs          10
## 394 hus_DTiKzwvw1Feny8xs                Dashboard Auth           2
## 395 hus_DTiKzwvw1Feny8xs             Dashboard Clients         119
## 396 hus_DTiKzwvw1Feny8xs          Dashboard Connections          34
## 397 hus_DTiKzwvw1Feny8xs          Dashboard Extensions           4
## 398 hus_DTiKzwvw1Feny8xs          Dashboard Login Page           1
## 399 hus_DTiKzwvw1Feny8xs                Dashboard Logs           7
## 400 hus_DTiKzwvw1Feny8xs                 Dashboard MFA           1
## 401 hus_DTiKzwvw1Feny8xs               Dashboard Rules           2
## 402 hus_DTiKzwvw1Feny8xs              Docs Connections           1
## 403 hus_DTiKzwvw1Feny8xs                  Docs General           7
## 404 hus_DTiKzwvw1Feny8xs                Docs Protocols          18
## 405 hus_DTiKzwvw1Feny8xs               Docs Quickstart          31
## 406 hus_DTiKzwvw1Feny8xs                        Home          34
## 407 hus_DTiKzwvw1Feny8xs                  Signup page           1
## 408 hus_eMafNQXMOJSm2fPN                        <NA>           0
## 409 hus_FtPZ9fNOikLNMjo6                        <NA>           0
## 410 hus_FtXMvSJDklY8ON3v                        <NA>           0
## 411 hus_HeJ8nmtqj9m9qwGT                        <NA>           0
## 412 hus_hUfFpFnqNDMDJeog                        Home           6
## 413 hus_hUfFpFnqNDMDJeog                     Pricing         613
## 414 hus_i7IP3dqS2SPQ2cOk                        <NA>           0
## 415 hus_iBlBXbQG7kzwW1Ez                        <NA>           0
## 416 hus_iJckQ078KuHB36p1                   Dashboard          27
## 417 hus_iJckQ078KuHB36p1           Dashboard 3rd Party           3
## 418 hus_iJckQ078KuHB36p1                Dashboard APIs          28
## 419 hus_iJckQ078KuHB36p1                Dashboard Auth           1
## 420 hus_iJckQ078KuHB36p1 Dashboard Breached Password           2
## 421 hus_iJckQ078KuHB36p1             Dashboard Clients          30
## 422 hus_iJckQ078KuHB36p1          Dashboard Connections          19
## 423 hus_iJckQ078KuHB36p1               Dashboard Emails           1
## 424 hus_iJckQ078KuHB36p1          Dashboard Extensions           2
## 425 hus_iJckQ078KuHB36p1               Dashboard Logins          14
## 426 hus_iJckQ078KuHB36p1                Dashboard Rules           2
```

```
## 427 hus_iJckQ078KuHB36p1                   Docs API           9
## 428 hus_iJckQ078KuHB36p1                Docs General           4
## 429 hus_iJckQ078KuHB36p1                    Docs IdP           1
## 430 hus_iJckQ078KuHB36p1                    Docs MFA           1
## 431 hus_iJckQ078KuHB36p1              Docs Quickstart          32
## 432 hus_iJckQ078KuHB36p1                        Home          39
## 433 hus_iJckQ078KuHB36p1                     Pricing         334
## 434 hus_iJckQ078KuHB36p1                  Signup page           1
## 435 hus_iJckQ078KuHB36p1                    Solutions           1
## 436 hus_iS01nQCPDZoVjdUC                         <NA>           0
## 437 hus_KnUSS5orPSNWEnpQ                        Blog           20
## 438 hus_KnUSS5orPSNWEnpQ                   Dashboard           41
## 439 hus_KnUSS5orPSNWEnpQ         Dashboard 3rd Party           3
## 440 hus_KnUSS5orPSNWEnpQ              Dashboard APIs           8
## 441 hus_KnUSS5orPSNWEnpQ Dashboard Breached Password           2
## 442 hus_KnUSS5orPSNWEnpQ            Dashboard Clients          22
## 443 hus_KnUSS5orPSNWEnpQ       Dashboard Connections          14
## 444 hus_KnUSS5orPSNWEnpQ        Dashboard Extensions           4
## 445 hus_KnUSS5orPSNWEnpQ        Dashboard Login Page           4
## 446 hus_KnUSS5orPSNWEnpQ            Dashboard Logins           20
## 447 hus_KnUSS5orPSNWEnpQ              Dashboard Logs           13
## 448 hus_KnUSS5orPSNWEnpQ               Dashboard MFA           5
## 449 hus_KnUSS5orPSNWEnpQ             Dashboard Rules           10
## 450 hus_KnUSS5orPSNWEnpQ           Docs hosted pages           1
## 451 hus_KnUSS5orPSNWEnpQ           Docs Library Lock           5
## 452 hus_KnUSS5orPSNWEnpQ             Docs Quickstart           4
## 453 hus_KnUSS5orPSNWEnpQ                  Docs Rules           3
## 454 hus_KnUSS5orPSNWEnpQ              Docs Tutorials           3
## 455 hus_KnUSS5orPSNWEnpQ                        Home          41
## 456 hus_KuGTmSF07m8S9WlH                   Dashboard           8
## 457 hus_KuGTmSF07m8S9WlH         Dashboard 3rd Party           1
## 458 hus_KuGTmSF07m8S9WlH              Dashboard APIs          12
## 459 hus_KuGTmSF07m8S9WlH            Dashboard Clients           6
## 460 hus_KuGTmSF07m8S9WlH                 Docs General           1
## 461 hus_KuGTmSF07m8S9WlH               Docs Quickstart           5
## 462 hus_KuGTmSF07m8S9WlH                       Support           2
## 463 hus_kUlq6L4RiDOqalHr                         <NA>           0
## 464 hus_lnlBsmn0xy2pRury                         <NA>           0
## 465 hus_MftgEMTYQO7rw283                         <NA>           0
## 466 hus_MRQmrRcWpoB4hORS                         <NA>           0
## 467 hus_MTw5WpOLId8zySmU                   Dashboard           41
## 468 hus_MTw5WpOLId8zySmU         Dashboard 3rd Party           1
## 469 hus_MTw5WpOLId8zySmU              Dashboard APIs           2
## 470 hus_MTw5WpOLId8zySmU              Dashboard Auth           2
## 471 hus_MTw5WpOLId8zySmU            Dashboard Clients          101
## 472 hus_MTw5WpOLId8zySmU       Dashboard Connections           4
## 473 hus_MTw5WpOLId8zySmU            Dashboard Logins           20
## 474 hus_MTw5WpOLId8zySmU            Docs Connections           2
## 475 hus_MTw5WpOLId8zySmU                 Docs Emails           1
## 476 hus_MTw5WpOLId8zySmU                Docs General           3
## 477 hus_MTw5WpOLId8zySmU                    Docs IdP           5
## 478 hus_MTw5WpOLId8zySmU              Docs Protocols           1
## 479 hus_MTw5WpOLId8zySmU              Docs Quickstart          16
## 480 hus_MTw5WpOLId8zySmU               Learn Tutorial           1
```

```
## 481 hus_MuVwnupCIP3EsPSz                    <NA>           0
## 482 hus_nHubVXDLBGgTvNcb                    <NA>           0
## 483 hus_ogb3hDPKAif3RxpV                    Blog          28
## 484 hus_ogb3hDPKAif3RxpV                Dashboard          40
## 485 hus_ogb3hDPKAif3RxpV      Dashboard 3rd Party           5
## 486 hus_ogb3hDPKAif3RxpV        Dashboard Account           9
## 487 hus_ogb3hDPKAif3RxpV           Dashboard APIs           6
## 488 hus_ogb3hDPKAif3RxpV           Dashboard Auth           3
## 489 hus_ogb3hDPKAif3RxpV        Dashboard Clients          67
## 490 hus_ogb3hDPKAif3RxpV    Dashboard Connections          17
## 491 hus_ogb3hDPKAif3RxpV     Dashboard Login Page           2
## 492 hus_ogb3hDPKAif3RxpV         Dashboard Logins          12
## 493 hus_ogb3hDPKAif3RxpV           Dashboard Logs           1
## 494 hus_ogb3hDPKAif3RxpV            Dashboard MFA           1
## 495 hus_ogb3hDPKAif3RxpV          Dashboard Rules           1
## 496 hus_ogb3hDPKAif3RxpV                 Docs API           2
## 497 hus_ogb3hDPKAif3RxpV       Docs Connections           3
## 498 hus_ogb3hDPKAif3RxpV           Docs General           7
## 499 hus_ogb3hDPKAif3RxpV         Docs Protocols           1
## 500 hus_ogb3hDPKAif3RxpV         Docs Quickstart          42
## 501 hus_ogb3hDPKAif3RxpV         Docs Tutorials           3
## 502 hus_ogb3hDPKAif3RxpV                   Home          11
## 503 hus_ogb3hDPKAif3RxpV              Signup page           1
## 504 hus_ogb3hDPKAif3RxpV                Solutions           1
## 505 hus_q4Gn4xzQOsxBZGHy                    <NA>           0
## 506 hus_qcVEdan61Fop67d6                    <NA>           0
## 507 hus_qsTNUNioVYAILa5t                    <NA>           0
## 508 hus_qt4kbCKYemfAXNmO                    <NA>           0
## 509 hus_qUePsKyOmOak3gts                    <NA>           0
## 510 hus_r4UHKtm5EQhEVKAE                    <NA>           0
## 511 hus_r5qV5b916rsbz6Uu                    <NA>           0
## 512 hus_RHMMMsjVSsMDVT6y                    <NA>           0
## 513 hus_s1u7FkL1I1S5BLOk                    <NA>           0
## 514 hus_s3doN3MHJ22sPLHk                    <NA>           0
## 515 hus_sPlMeFjSSuKmsVfI                Dashboard          39
## 516 hus_sPlMeFjSSuKmsVfI      Dashboard 3rd Party           1
## 517 hus_sPlMeFjSSuKmsVfI           Dashboard APIs          10
## 518 hus_sPlMeFjSSuKmsVfI           Dashboard Auth           1
## 519 hus_sPlMeFjSSuKmsVfI        Dashboard Clients          48
## 520 hus_sPlMeFjSSuKmsVfI    Dashboard Connections          12
## 521 hus_sPlMeFjSSuKmsVfI         Dashboard Emails           1
## 522 hus_sPlMeFjSSuKmsVfI     Dashboard Login Page          10
## 523 hus_sPlMeFjSSuKmsVfI         Dashboard Logins          14
## 524 hus_sPlMeFjSSuKmsVfI           Dashboard Logs           6
## 525 hus_sPlMeFjSSuKmsVfI            Dashboard MFA           8
## 526 hus_sPlMeFjSSuKmsVfI          Dashboard Rules           4
## 527 hus_sPlMeFjSSuKmsVfI           Docs General           2
## 528 hus_sPlMeFjSSuKmsVfI         Docs Quickstart          18
## 529 hus_sPlMeFjSSuKmsVfI              Signup page           1
## 530 hus_sPlMeFjSSuKmsVfI                  Support           2
## 531 hus_TOLtMdOVIfFFLG3K                    <NA>           0
## 532 hus_tbT8bQkYAY7vC8aC                    <NA>           0
## 533 hus_TJRCm4JyKTPgZLGp                    <NA>           0
## 534 hus_TmsLeDgw8O4GBH2H                    <NA>           0
```

```
## 535 hus_TNN3uQJI9qYXAYyP                       <NA>           0
## 536 hus_tQQlfP8EIJZ8x12U                       <NA>           0
## 537 hus_w5qVGHn96mXRvaHF                  Dashboard          25
## 538 hus_w5qVGHn96mXRvaHF          Dashboard 3rd Party           5
## 539 hus_w5qVGHn96mXRvaHF            Dashboard Account          23
## 540 hus_w5qVGHn96mXRvaHF               Dashboard APIs          18
## 541 hus_w5qVGHn96mXRvaHF               Dashboard Auth           1
## 542 hus_w5qVGHn96mXRvaHF Dashboard Breached Password          13
## 543 hus_w5qVGHn96mXRvaHF            Dashboard Clients          27
## 544 hus_w5qVGHn96mXRvaHF        Dashboard Connections          23
## 545 hus_w5qVGHn96mXRvaHF             Dashboard Emails           5
## 546 hus_w5qVGHn96mXRvaHF         Dashboard Extensions           3
## 547 hus_w5qVGHn96mXRvaHF         Dashboard Login Page          12
## 548 hus_w5qVGHn96mXRvaHF             Dashboard Logins          49
## 549 hus_w5qVGHn96mXRvaHF               Dashboard Logs           2
## 550 hus_w5qVGHn96mXRvaHF                Dashboard MFA           5
## 551 hus_w5qVGHn96mXRvaHF              Dashboard Rules          14
## 552 hus_w5qVGHn96mXRvaHF                     Docs API           1
## 553 hus_w5qVGHn96mXRvaHF             Docs Connections           1
## 554 hus_w5qVGHn96mXRvaHF                 Docs General           2
## 555 hus_w5qVGHn96mXRvaHF            Docs hosted pages           4
## 556 hus_w5qVGHn96mXRvaHF              Docs Quickstart          26
## 557 hus_w5qVGHn96mXRvaHF                        Home          49
## 558 hus_w5qVGHn96mXRvaHF                     Pricing           1
## 559 hus_wceN4FvRwJAC1Skp                       <NA>           0
## 560 hus_wKONGXNI1cORRNtE                       <NA>           0
## 561 hus_x4itnv7S3qEoYNrN                       <NA>           0
## 562 hus_XgyOQtLl4WOMiBvw                       <NA>           0
## 563 hus_XjXWfi6Xy7PsJIMv                       <NA>           0
## 564 hus_YP9UBIkrewcZYl3U                       <NA>           0
## 565 hus_yRhhBnL9n3qv42BW                       <NA>           0
## 566 hus_YSWac6z9DB6kGAEz                       <NA>           0
## 567 hus_zIrajKH2AL5S3qeL                       <NA>           0
## 568 hus_ZQo19eYi4CZ253z7                  Dashboard           8
## 569 hus_ZQo19eYi4CZ253z7          Dashboard 3rd Party           1
## 570 hus_ZQo19eYi4CZ253z7               Dashboard APIs           6
## 571 hus_ZQo19eYi4CZ253z7            Dashboard Clients          16
## 572 hus_ZQo19eYi4CZ253z7        Dashboard Connections           4
## 573 hus_ZQo19eYi4CZ253z7             Dashboard Emails           2
## 574 hus_ZQo19eYi4CZ253z7             Dashboard Logins           9
## 575 hus_ZQo19eYi4CZ253z7              Dashboard Rules           2
## 576 hus_ZQo19eYi4CZ253z7                     Docs API           1
## 577 hus_ZQo19eYi4CZ253z7                 Docs General           1
## 578 hus_ZQo19eYi4CZ253z7              Docs Quickstart          12
## 579 hus_ZQo19eYi4CZ253z7               Docs Tutorials           1
## 580 hus_ZQo19eYi4CZ253z7                        Home          18
## 581             linkedin                       About           4
## 582             linkedin                   AuthCatalog         159
## 583             linkedin                  Availability           1
## 584             linkedin                          Blog         249
## 585             linkedin                     Dashboard        3456
## 586             linkedin          Dashboard 3rd Party         127
## 587             linkedin            Dashboard Account         247
## 588             linkedin               Dashboard APIs         897
```

```
## 589            linkedin              Dashboard Auth         437
## 590            linkedin Dashboard Breached Password          20
## 591            linkedin           Dashboard Clients        1679
## 592            linkedin       Dashboard Connections        1213
## 593            linkedin            Dashboard Emails          13
## 594            linkedin        Dashboard Extensions         127
## 595            linkedin            Dashboard invite          33
## 596            linkedin        Dashboard Login Page          85
## 597            linkedin            Dashboard Logins        1645
## 598            linkedin              Dashboard Logs         189
## 599            linkedin               Dashboard MFA          84
## 600            linkedin             Dashboard Rules         275
## 601            linkedin                    Docs API         328
## 602            linkedin             Docs Appliance          12
## 603            linkedin           Docs Connections         611
## 604            linkedin             Docs Extensions          23
## 605            linkedin                Docs General         282
## 606            linkedin          Docs hosted pages          14
## 607            linkedin                    Docs IdP          56
## 608            linkedin        Docs Library Auth0js          44
## 609            linkedin           Docs Library Lock         110
## 610            linkedin   Docs Library Lock Android          26
## 611            linkedin       Docs Library Lock iOS           3
## 612            linkedin                    Docs MFA           8
## 613            linkedin             Docs Protocols          22
## 614            linkedin            Docs Quickstart         946
## 615            linkedin                 Docs Rules          15
## 616            linkedin             Docs Scenarios          48
## 617            linkedin               Docs Support           1
## 618            linkedin             Docs Tutorials          20
## 619            linkedin                    E-Books         139
## 620            linkedin                     Hiring          47
## 621            linkedin                       Home        5396
## 622            linkedin                 Learn Guide          20
## 623            linkedin             Learn Tutorial           9
## 624            linkedin             Learn Use Case           2
## 625            linkedin                      Other           5
## 626            linkedin                    Pricing         987
## 627            linkedin                   Products          16
## 628            linkedin                  Resources           3
## 629            linkedin                      Rules           1
## 630            linkedin                   Security           1
## 631            linkedin                 Signup page          20
## 632            linkedin                  Solutions           6
## 633            linkedin                    Support         419
## 634            linkedin          Website Education          32
## 635            linkedin                       <NA>           0
## 636                waad                  Dashboard          20
## 637                waad        Dashboard 3rd Party           1
## 638                waad          Dashboard Account           5
## 639                waad             Dashboard Auth           2
## 640                waad                       Home          24
## 641         windowslive                      About          20
## 642         windowslive                AuthCatalog         475
```

```
## 643         windowslive                Availability          10
## 644         windowslive                        Blog        1194
## 645         windowslive                  Compliance           8
## 646         windowslive                   Dashboard       12372
## 647         windowslive          Dashboard 3rd Party        1415
## 648         windowslive            Dashboard Account        1212
## 649         windowslive               Dashboard APIs        7727
## 650         windowslive               Dashboard Auth         754
## 651         windowslive Dashboard Breached Password         162
## 652         windowslive            Dashboard Clients       13601
## 653         windowslive        Dashboard Connections        2781
## 654         windowslive             Dashboard Emails         468
## 655         windowslive         Dashboard Extensions         618
## 656         windowslive             Dashboard invite         164
## 657         windowslive         Dashboard Login Page         775
## 658         windowslive             Dashboard Logins       15501
## 659         windowslive               Dashboard Logs         796
## 660         windowslive                Dashboard MFA         430
## 661         windowslive              Dashboard Rules        3599
## 662         windowslive                  Docs Addons          15
## 663         windowslive                     Docs API        1583
## 664         windowslive                Docs Appliance          30
## 665         windowslive                     Docs CMS          19
## 666         windowslive              Docs Connections         709
## 667         windowslive                Docs Connector          10
## 668         windowslive                   Docs Emails          57
## 669         windowslive              Docs Extensions          85
## 670         windowslive                 Docs General        1603
## 671         windowslive            Docs hosted pages          79
## 672         windowslive                     Docs i18n           1
## 673         windowslive                     Docs IdP         234
## 674         windowslive            Docs Integrations          45
## 675         windowslive         Docs Library Auth0js         169
## 676         windowslive            Docs Library Lock         446
## 677         windowslive    Docs Library Lock Android          42
## 678         windowslive        Docs Library Lock iOS          35
## 679         windowslive                     Docs MFA          81
## 680         windowslive               Docs Protocols         123
## 681         windowslive              Docs Quickstart        5389
## 682         windowslive                   Docs Rules         126
## 683         windowslive               Docs Scenarios         105
## 684         windowslive                Docs Services           2
## 685         windowslive                Docs Support          18
## 686         windowslive               Docs Tutorials         165
## 687         windowslive                      E-Books         113
## 688         windowslive                       Hiring          20
## 689         windowslive                         Home       29754
## 690         windowslive             Learn Case Study           7
## 691         windowslive            Learn Competition           3
## 692         windowslive                  Learn Guide          28
## 693         windowslive               Learn Tutorial         112
## 694         windowslive               Learn Use Case          22
## 695         windowslive                   OpenSource          14
## 696         windowslive                        Other          14
```

```
## 697          windowslive                   Pricing        20333
## 698          windowslive                  Products          117
## 699          windowslive                 Resources            4
## 700          windowslive                     Rules            7
## 701          windowslive                  Security           25
## 702          windowslive                Signup page          324
## 703          windowslive                 Solutions           44
## 704          windowslive                   Support         1068
## 705          windowslive                University           71
## 706          windowslive              User Program            4
## 707          windowslive           Website Education           94
## 708          windowslive                      <NA>            0
```

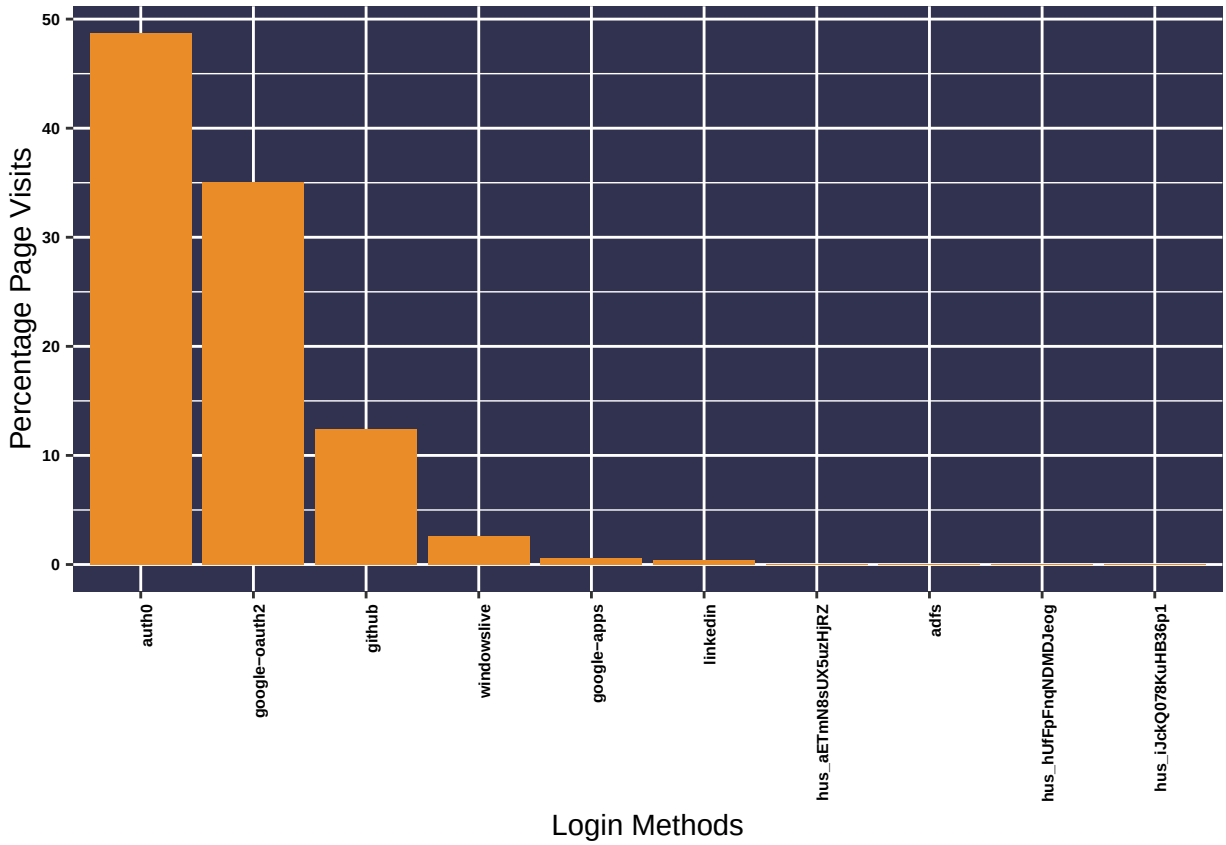
```
# Identifying the distribution of page views for each login method without considering the individual p
page_views_just_with_login_method <- dataset_emails %>%
  left_join(dataset_pageviews, by ="email_id") %>%
  mutate(visits = ifelse(is.na(visits),0,visits))%>%
  select(login_method, visits) %>%
  group_by(login_method) %>%
  summarize(total_visits = sum(visits)) %>%
  top_n(10)
```

```
## Selecting by total_visits
```

```
print(data.frame(page_views_just_with_login_method))
```

```
##            login_method total_visits
## 1                  adfs         1576
## 2                 auth0      2341139
## 3                github       595127
## 4           google-apps        30018
## 5         google-oauth2      1683006
## 6  hus_aETmN8sUX5uzHjRZ         3271
## 7  hus_hUfFpFnqNDMDJeog          619
## 8  hus_iJckQ078KuHB36p1          551
## 9               linkedin        20612
## 10          windowslive       127431
```

```
# Finding the percentage views and ordering them
page_views_just_with_login_method$total_visits <- page_views_just_with_login_method$total_visits  / sum
page_views_just_with_login_method$login_method <- factor(page_views_just_with_login_method$login_method

# Creating a plot of page views per login method
p4 <- ggplot() + geom_bar(aes(y = total_visits, x = login_method), data = page_views_just_with_login_me
                          stat = "identity", fill="#ea8c27") +
  theme(panel.background = element_rect(fill = "#31324F",
                                colour = "#31324F",
                                size = 0.5, linetype = "solid"),  axis.text.x = element_text(face="bold
        axis.text.y = element_text(face="bold", color="black", size=6)) +ylab("Percentage Page Visits")

# printing the distribution of page views per login method
p4
```



Refering to visualization and dataframe - "page_views_just_with_login_method"

As you can see in the sorted plot, the distribution is mostly divided into six login_methods as we can see the significant percentage of page_visits for these login methods - auth0 , google-oauth2, github, windowslive, google-apps and linkedin. The login methods like "hus_iJckQ078KuHB36p1" looks unfamiliar to me and I would definitely consider doing more research as it might be the case of unauthorized login and thats why the data appears like this.

Question 5

What differences do you find between developers and non-developers?

```
# Filtering emails_dataset to have developers and non-developers separated without considering blank ro
role_df <- dataset_emails %>%
  filter(role!="") %>%
  mutate(Designation = ifelse(role %in% c("developer", "Software Developer", "Software Engineer III, Wo
                                          Analytics"),"Developer","Non-Developer"))
```

Considering the Numeric Variables

Identifying the numeric data variables to see if there are any significant differences between developers and non developers by considering the mean values for different numeric data points w.r.t

```r
# Mean & Total page visits by developer and non developers
role_page_visits <- role_df %>%
  left_join(dataset_pageviews,by="email_id") %>%
  mutate(visits = ifelse(is.na(visits),0,visits)) %>%
  group_by(Designation) %>%
  summarize(avg=mean(visits))

role_page_visits <- spread(role_page_visits, Designation, avg)
rownames(role_page_visits) <- "Page Visits"
```

```
## Warning: Setting row names on a tibble is deprecated.
```

```r
# Mean & Total "account_total_tickets" by developer vs non developer
tickets_total <- role_df %>%
  left_join(dataset_relation, by="email_id") %>%
  left_join(dataset_tenants, by="tenant_id") %>%
  mutate(account_total_tickets = ifelse(is.na(account_total_tickets),0,account_total_tickets)) %>%
  group_by(Designation) %>%
  summarize(avg=mean(account_total_tickets))

tickets_total <- spread(tickets_total, Designation, avg)
rownames(tickets_total) <- "account_total_tickets"
```

```
## Warning: Setting row names on a tibble is deprecated.
```

```r
# Mean & Total "account_open_tickets" by developer vs non developer
tickets_open <- role_df %>%
  left_join(dataset_relation, by="email_id") %>%
  left_join(dataset_tenants, by="tenant_id") %>%
  mutate(account_open_tickets = ifelse(is.na(account_open_tickets),0,account_open_tickets)) %>%
  group_by(Designation) %>%
  summarize(avg=mean(account_open_tickets))

tickets_open <- spread(tickets_open, Designation, avg)
rownames(tickets_open) <- "account_open_tickets"
```

```
## Warning: Setting row names on a tibble is deprecated.
```

```r
# Mean & Total Dashboard Sessions by developer vs non developer
sessions_dashboard <- role_df %>%
  mutate(dashboard_sessions = ifelse(is.na(dashboard_sessions),0,dashboard_sessions)) %>%
  group_by(Designation) %>%
  summarize(avg=mean(dashboard_sessions))

sessions_dashboard <- spread(sessions_dashboard, Designation, avg)
rownames(sessions_dashboard) <- "Average Dashboard Sessions"
```

```
## Warning: Setting row names on a tibble is deprecated.
```

```r
# Mean & Total active users last 30 days for developers and non developers
role_df_active_users_30d <- role_df %>%
  left_join(dataset_relation,by="email_id") %>%
  left_join(dataset_tenants, by="tenant_id") %>%
  mutate(active_users_last_30d = ifelse(is.na(active_users_last_30d),0,active_users_last_30d)) %>%
  group_by(Designation) %>%
  summarize(avg=mean(active_users_last_30d))

role_df_active_users_30d <- spread(role_df_active_users_30d, Designation, avg)
rownames(role_df_active_users_30d) <- "Average Number of active users in last 30 days"
```

```
## Warning: Setting row names on a tibble is deprecated.
```

```r
# Mean & Total "total apps" for developers and non developers
role_df_total_apps_tenants <- role_df %>%
  left_join(dataset_relation,by="email_id") %>%
  left_join(dataset_tenants, by="tenant_id") %>%
  mutate(total_apps = ifelse(is.na(total_apps),0,total_apps)) %>%
  group_by(Designation) %>%
  summarize(avg=mean(total_apps))

role_df_total_apps_tenants <- spread(role_df_total_apps_tenants, Designation, avg)
rownames(role_df_total_apps_tenants)<-"Average Number of Applications configured per tenant"
```

```
## Warning: Setting row names on a tibble is deprecated.
```

Pre-processing for categorical variables

```r
# Joining role_df with tenants datframe using relations dataframe.
role_df_tenants_combined<- role_df%>% left_join(dataset_relation,by="email_id") %>%  left_join(dataset_

#calculate_percentage calculates the percentage value for every column value
calculate_percentage <- function(dataframe){
dataframe$Developer <- dataframe$Developer  / sum(dataframe$Developer)*100
dataframe$`Non-Developer`  <- dataframe$`Non-Developer`   / sum(dataframe$`Non-Developer` )*100
return(dataframe)
}

# Trim function using gsub() to remove all the leading and trailing whitespaces
trim <- function (x) gsub("^\\s+|\\s+$", "", x)
```

Categorical Variables analysis

```r
# Login Method count differences between developer and non developer
role_df_login <- role_df_tenants_combined %>%
  group_by(Designation, login_method) %>%
  summarise(count=n())
```

```
role_df_login <- spread(role_df_login, Designation, count) %>% top_n(5)
```

Selecting by Non-Developer

```
role_df_login <- calculate_percentage(role_df_login)

# Source Group count differences between developer and non developer
role_df_source_group <- role_df_tenants_combined %>%
  group_by(Designation, source_group) %>%
  summarise(count=n())

role_df_source_group <- spread(role_df_source_group, Designation, count) %>% top_n(5)
```

Selecting by Non-Developer

```
role_df_source_group <- calculate_percentage(role_df_source_group)

# Operating System count differences between developer and non developer
role_df_operating_system <- role_df_tenants_combined %>%
  group_by(Designation, sperating_system) %>%
  summarise(count=n())

role_df_operating_system <- spread(role_df_operating_system, Designation, count) %>% top_n(5)
```

Selecting by Non-Developer

```
role_df_operating_system <- calculate_percentage(role_df_operating_system)

# Analyzing the technologies used by developers and non developers
s <- strsplit(role_df_tenants_combined$technologies_used, split = ",")
tenants_technologies_combined <- data.frame(tenant_id = rep(role_df_tenants_combined$tenant_id, sapply(s
tenants_technologies_combined$technologies_used <- trim(tenants_technologies_combined$technologies_used)
tenants_technologies_combined = as.data.frame(lapply(tenants_technologies_combined, na.omit))

technologies_compare <- role_df_tenants_combined[,c("Designation", "tenant_id")] %>% left_join(tenants_t
  select(Designation, technologies_used) %>%
  group_by(Designation, technologies_used) %>%
  summarise(count=n())
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factor and character vector, coercing into character vector
```

```
technologies_compare  <- spread(technologies_compare , Designation, count) %>% na.omit()
technologies_compare <- calculate_percentage(technologies_compare)

# Analyzing the connection types of developers and non developers
s <- strsplit(role_df_tenants_combined$connection_types, split = ",")
tenants_connections_combined <- data.frame(tenant_id = rep(role_df_tenants_combined$tenant_id, sapply(s
tenants_connections_combined$connection_types <- trim(tenants_connections_combined$connection_types)
tenants_connections_combined = as.data.frame(lapply(tenants_connections_combined, na.omit))
```

```r
connections_compare <- role_df_tenants_combined[,c("Designation", "tenant_id")] %>% left_join(tenants_c
  select(Designation, connection_types) %>%
  group_by(Designation, connection_types) %>%
  summarise(count=n())
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factor and character vector, coercing into character vector
```

```r
connections_compare  <- spread(connections_compare , Designation, count)
connections_compare <- calculate_percentage(connections_compare)

# Analyzing the features used by developers and non developers
s <- strsplit(role_df_tenants_combined$used_features, split = ",")
tenants_features_combined <- data.frame(tenant_id = rep(role_df_tenants_combined$tenant_id, sapply(s, l
tenants_features_combined$used_features <- trim(tenants_features_combined$used_features)

tenants_features_combined <- as.data.frame(lapply(tenants_features_combined, na.omit))
features_compare <- role_df_tenants_combined[,c("Designation", "tenant_id")] %>% left_join(tenants_featu
  select(Designation, used_features) %>%
  group_by(Designation, used_features) %>%
  summarise(count=n())
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factor and character vector, coercing into character vector
```

```r
features_compare  <- spread(features_compare , Designation, count)
features_compare <- calculate_percentage(features_compare)

# Combibining all the numeric averages in a single table to see the differences between developers and

tickets_total <- data.frame(tickets_total)
tickets_open <- data.frame(tickets_open)
sessions_dashboard <- data.frame(sessions_dashboard)
role_df_active_users_30d <- data.frame(role_df_active_users_30d)
role_df_total_apps_tenants <- data.frame(role_df_total_apps_tenants)

# Using Row Bind to construct a single comparison dataframe
comparisons <- rbind(tickets_total, tickets_open, sessions_dashboard, role_df_active_users_30d, role_df
colnames(comparisons)
```

```
## [1] "Developer"     "Non.Developer"
```

```r
ncol(comparisons)
```

```
## [1] 2
```

```r
comparisons$Entity <- row.names(comparisons)
rownames(comparisons) <- NULL

#Dispaying comparisons of numeric variables
print(comparisons)
```

```
##     Developer Non.Developer
## 1  0.51308494     2.4528536
## 2  0.03836771     0.1600496
## 3  8.11574347    10.5988740
## 4 23.32468397  2384.1557072
## 5  2.28609448     7.2884615
##                                                  Entity
## 1                                 account_total_tickets
## 2                                  account_open_tickets
## 3                               Average Dashboard Sessions
## 4       Average Number of active users in last 30 days
## 5 Average Number of Applications configured per tenant
```

```
# Displaying the result of the cateogrical variables
role_df_login <- data.frame(role_df_login)
print(role_df_login)
```

```
##    login_method Developer Non.Developer
## 1         auth0 39.100924     43.063402
## 2        github 17.848003      7.470182
## 3 google-oauth2 38.978525     44.569994
## 4      linkedin  1.268499      2.259887
## 5   windowslive  2.804050      2.636535
```

```
role_df_source_group <- data.frame(role_df_source_group)
print(role_df_source_group)
```

```
##   source_group Developer Non.Developer
## 1  Auth0 Child  6.029339      3.531856
## 2       Direct 19.579811     35.041551
## 3       Google 67.391845     49.307479
## 4  Metrics Bug  1.864744      2.354571
## 5        Other  5.134262      9.764543
```

```
role_df_operating_system <- data.frame(role_df_operating_system)
print(role_df_operating_system)
```

```
##   sperating_system Developer Non.Developer
## 1          Android  1.751208      4.726027
## 2            Linux  9.371981      4.178082
## 3         Mac OS X 40.096618     47.739726
## 4       Windows 10 36.533816     30.479452
## 5        Windows 7 12.246377     12.876712
```

```
technologies_compare <- data.frame(technologies_compare)
print(technologies_compare)
```

```
##    technologies_used   Developer Non.Developer
## 1            android  1.61123735     1.9292605
## 2              apple  1.71452179     2.7560864
## 3           auth0_js 10.41107209    22.4161690
```

```
## 4             dotnet  0.74364801     2.2967386
## 5      dotnet_manage  1.11547201     4.3638034
## 6               java  4.58582937    10.8406063
## 7            laravel  0.78496178     0.5052825
## 8               lock 70.50196240    40.0091870
## 9  lock-passwordless  0.86758934     0.9646302
## 10          omniauth  0.06197067     0.2296739
## 11              owin  1.44598224     2.8479559
## 12               php  1.63189424     3.3073036
## 13     react_native  0.86758934     0.1837391
## 14              ruby  0.92956001     2.8479559
## 15         wordpress  2.72670936     4.5016077
```

```
features_compare <- data.frame(features_compare)
print(features_compare)
```

```
##            used_features  Developer Non.Developer
## 1 custom email providers  0.7360901      7.263682
## 2                 emails  0.9742368      8.109453
## 3          impersonation  0.6819658      5.422886
## 4  password policy setup  0.6819658      4.875622
## 5                   <NA> 96.9257415     74.328358
```

```
connections_compare <- data.frame(connections_compare)
print(connections_compare)
```

```
##   connection_types Developer Non.Developer
## 1         customDB  2.200560      8.948247
## 2       enterprise  1.679375      7.245409
## 3       passwordle  1.785542      2.570952
## 4           social 14.593186     18.731219
## 5      user & pass 11.475726     27.946578
## 6             <NA> 68.265611     34.557596
```

Explanation for Numeric Variables Analysis: (using "comparisons" dataframe)

As you can see in the comparisons dataframe where the tradeoffs between numeric variables are identified based on the average value, some of the major differences identified between developers and non developers were -

1. Tickets:

The average number of open tickets for developers are 0.5 and for non-developers are 2.5 which is five times more than that of the developers. It might signifies that the people who are classified as non developers can be considered as less technical when it comes to system usage and hence, more tickets issued from the on developer group. Also, the average number of open tickets for non-developers 0.16 compared to developers which is 0.04, a four times more which might imply that more time is taken to handle the tickets of non developer group compared to developer group. Analyzing the tickets can be considered an important task for future analysis as the ticket solving takes time and resources, so understanding the factors from wider perspective can be beneficial to reduce the total number of tickets issued by both groups.

2. Acpplications Configured:

The average number of applications configured per tenant for developer group is around 2.3 compared to non-developer group which is 7.3 which shows the higher ratio (more than 3 times) for non developer group

3. Average Number of Active Users:

The average number of active users in the last 30 days in non-developer group is 2385 which is 100 times more compared to developer group, for hwich it is aorund 23, which shows users in non-developer are on avergae way more active compared to developer group users average for the month of July.

Explanation for Categorical Variables Analysis: (using "comparisons" dataframe)

1. Connection_type ("conneciton_compare" dataframe):

While most of the entries for connection type were blank, one of the significant different that was noiticed was with enterprise connection type, where there was a signifacant differece in terms of the proportion. Non-developer showed higher number of overall enterprise conenction compared to developer averages.

2. Features Use ("feature_compare" dataframe):

For the available datapoints in terms of the features in all four categories (custom_email_providers, emails, impersonation, password_policy_setup), non-developer group showed on an average 8 times more number of users using the features in all four categories compared to non-developers, which shows that non-developer group is more inclined towards the features provided by Auth0 products compared to developer group.

3. Technologies Used ("technologies_compare" dataframe):

Although there were similar trends in terms of the different technologies used by both developers and non-developer group, the significant difference can be observed when we see how the average number of people using each technology is distributed. In case of developer, we can see around 70% of the developers are using "lock" which is a high number and shows the importance of "lock" in terms of adding new fetures for developers as well as for their continued engagement. Non-developer group showed highest numbers for "lock" (40%), "auth0_js"(22%) & "Java"(10%) thus showing the less skewed distribution unlike developer group in terms of the technologies.

Question 6

What are the most and least used technologies in tenants?

Based on the results below, "lock is the most used technology" while "node_auth0" is the least used technology

```
# Splitting the rows for the column values in "technologies_used" to count for the multiple technologie
s <- strsplit(dataset_tenants$technologies_used, split = ",")

tenants_technologies <- data.frame(tenant_id = rep(dataset_tenants$tenant_id, sapply(s, length)), techno

tenants_technologies$technologies_used <- trim(tenants_technologies$technologies_used)

# Most used technology in tenants
tenants_technologies_most <- tenants_technologies %>%
  select(technologies_used, tenant_id) %>%
```

```
  group_by(technologies_used) %>%
  summarise(count=n()) %>%
  filter(count==max(count))

# Least used technologies in tenants
tenants_technologies_least <- tenants_technologies %>%
  select(technologies_used, tenant_id) %>%
  group_by(technologies_used) %>%
  summarise(count=n()) %>%
  filter(count==min(count))

tenant_technologies_max_min = rbind(tenants_technologies_most, tenants_technologies_least)
print(data.frame(tenant_technologies_max_min))
```

```
##   technologies_used count
## 1              lock  4094
## 2       node_auth0     5
```

Question 7

Please mark any inconsistencies you find in the data that you'd research further if you were working at Auth0

During this analysis task, I come across a large number of missing values in different variables that can be consider as a main factor to make the data inconsistent.

1. Number of Paid Users per source group:

According to my understanding of the source groups, having an accurate data regarding the paid and unpaid users is highly important to identify the distribution of paid vs unpaid users within each source group. It can be helpful to further analyze the factors such as marketting strategies for the paid / unpaid users for each source group, behaviour of the paid vs unpaid users with respect to each page type, what is the total amount of time each user is spending on various pages and how all that might affect the revenue model of the product in ong run. I identified that there were several groups with missing values for the paid user column w.r.t these source groups that might make the analysis biased.

2. Connection Types:

During the analysis of the developers and non developers differences, there were several missing values for the connections. To be more specific, developers column had 68% (7074) missing values and non-developers column had 35% (1034) of missng values which is a significant amount of missin data for the important technical entity like connection_type.

```
# Prints the number of connection types for developers and non developers
s <- strsplit(role_df_tenants_combined$connection_types, split = ",")
tenants_connections_combined <- data.frame(tenant_id = rep(role_df_tenants_combined$tenant_id, sapply(s
tenants_connections_combined$connection_types <- trim(tenants_connections_combined$connection_types)
tenants_connections_combined = as.data.frame(lapply(tenants_connections_combined, na.omit))
connections_compare_inc <- role_df_tenants_combined[,c("Designation", "tenant_id")] %>% left_join(tenan
  select(Designation, connection_types) %>%
  group_by(Designation, connection_types) %>%
  summarise(count=n())
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factor and character vector, coercing into character vector
```

```
connections_compare_inc  <- spread(connections_compare_inc , Designation, count)
print(connections_compare_inc)
```

```
## # A tibble: 6 × 3
##   connection_types Developer `Non-Developer`
## *           <fctr>     <int>           <int>
## 1         customDB       228             268
## 2       enterprise       174             217
## 3       passwordle       185              77
## 4           social      1512             561
## 5      user & pass      1189             837
## 6               NA      7073            1035
```

3. Designation:

Designation(role) was missing from 3239 out of 13,000 entries (25%) in the dataset_emails column, which is a significant number and it might reduce the accuracy when it comes to finalizing the decisions where the role of the tenant plays an important part. One of the reasons could be not markeing the field compulsary when taking input from users in UI. Making it a mandatory option during registration might help reducing such inconsistencies from gathering the data.

```
# printing the number of missing (blank) values for the role(designation)
role_blanks <- dataset_emails %>%
  subset(role == "") %>%
  summarise(count = n())

# Number of unidentified designations
print(role_blanks)
```

```
## # A tibble: 1 × 1
##   count
##   <int>
## 1  3239
```

4. Used Features Data points:

WHen considered for the use of individual features from the group (custom_email_providers, emails, impersonation & password policy setup), a huge amount of datapoijts were missing during the developers and non-developers difference analysis. 96% of the values from the developers group (around 7000) and 74% of the values from the non-developers (around 1200) groups were missing which shows a serious breach in the data collection strategy in terms of the feature use for the product. Having the data for each features w.r.t. several available data points would lead to highly insightful information when it comes to reviews of different features, modiciation strategies as well as addition of new features w.r.t current feature use.

5. Region:

```
print(unique(dataset_tenants$region))
```

```
## [1] "europe"    "master"    "australia" "pus2"
```

As we can see in the above code output, the resion field is not divided as per the continent, except for europe and australia. There are major number of users in United States, India and it is not divided based on the continent. This might be the internal representation based on the Auth0 product region mapping so I might be wrong here!

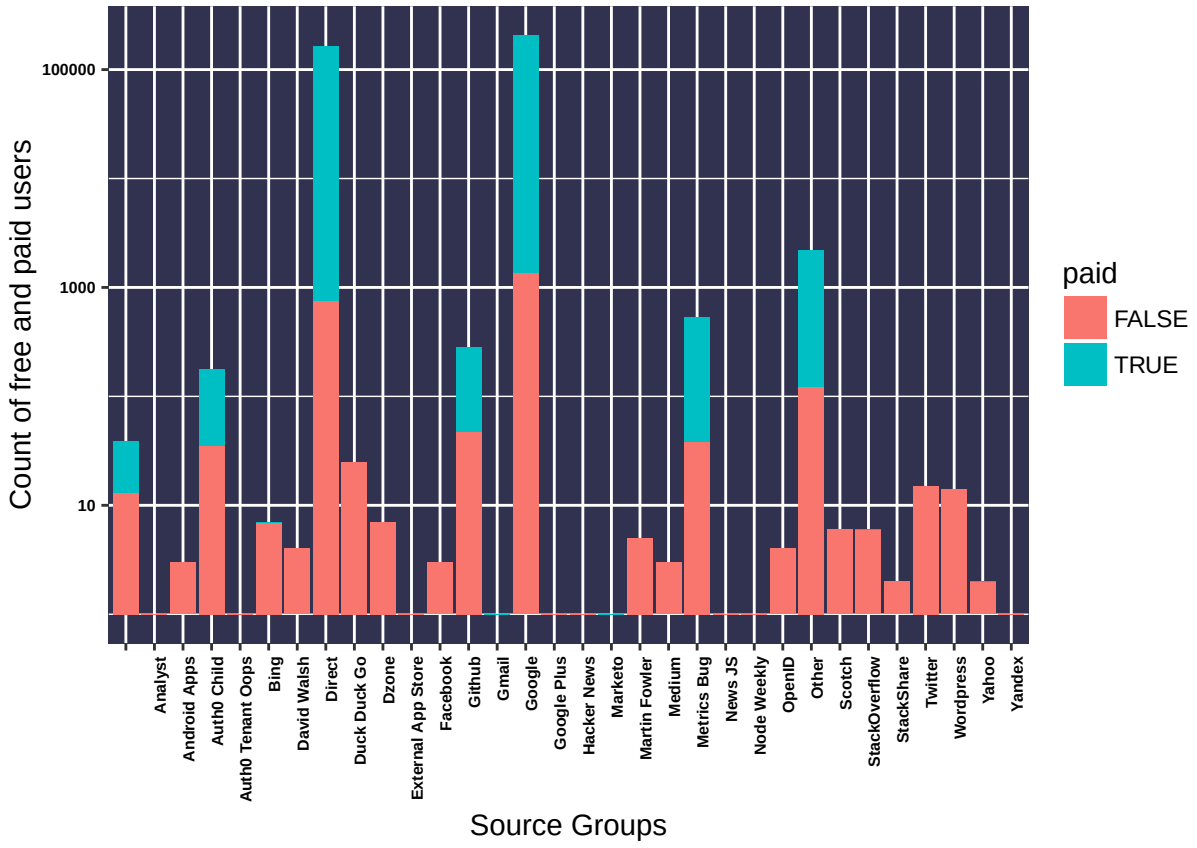
Question 8

Create at least 3 visualizations using R which you consider interesting for the provided data

1. Visualization of the source group w.r.t paid and unpaid users

Source group might be an important factor in terms of the monetization, prividing insights regarding how many paid users are there compared to number of free users for each group. It will further help directing the various domains such as marketing (to identify where to spend most money to market the paid features compared to groups using the free version of the product). I have created visualizations based on both count and percentage of paid vs unpaid users in each group for clearer understanding.

```
v <- dataset_emails %>%
  left_join(dataset_relation,by="email_id") %>%
  left_join(dataset_tenants,by="tenant_id") %>%
  group_by(source_group, paid) %>% na.omit() %>%
  summarize(count=n())

p4 <- ggplot() + geom_bar(aes(y = count, x = source_group, fill=paid ), data=v,
                          stat = "identity") +
  theme(panel.background = element_rect(fill = "#31324F",
                                colour = "#31324F",
                                size = 0.5, linetype = "solid"),  axis.text.x = element_text(face="bold
        axis.text.y = element_text(face="bold", color="black", size=6))+scale_y_log10() + ylab("Count o
print(p4)
```




```
# Using spread to reshape the dataframe
v1 <- spread(v, paid, count)
colnames(v1)[2] <- "free_users"
colnames(v1)[3] <- "paid_users"

# I have filtered the columns to have just the numeric values and then calculate the average for every
v1 <- data.frame(v1)
str(v1)
```

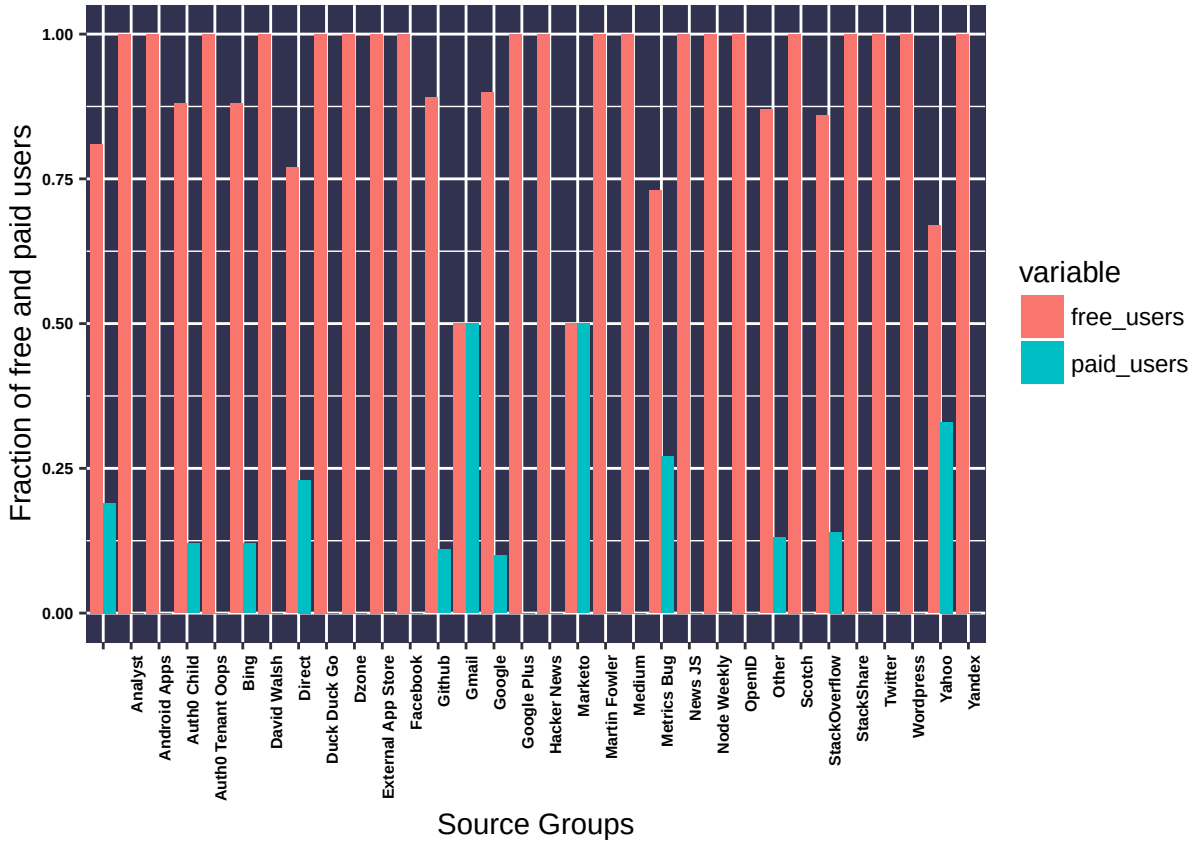
```
## 'data.frame':    32 obs. of  3 variables:
##  $ source_group: chr  "" "Analyst" "Android Apps" "Auth0 Child" ...
##  $ free_users  : int  13 1 3 35 1 7 4 751 25 7 ...
##  $ paid_users  : int  3 NA NA 5 NA 1 NA 219 NA NA ...
```

```
v1 <- v1 %>% mutate(paid_users = ifelse(is.na(paid_users),0,paid_users))
str(v1)
```

```
## 'data.frame':    32 obs. of  3 variables:
##  $ source_group: chr  "" "Analyst" "Android Apps" "Auth0 Child" ...
##  $ free_users  : int  13 1 3 35 1 7 4 751 25 7 ...
##  $ paid_users  : num  3 0 0 5 0 1 0 219 0 0 ...
```

```
v1$free_users <- as.numeric(v1$free_users)
v1$paid_users <- as.numeric(v1$paid_users)
v3 <- subset(v1, select=source_group)
```

```
v2 <- subset(v1, select=free_users:paid_users)
v2 <- round(prop.table(as.matrix(v2),1),2)
# Using column binding to construct the new dataframe with percentage users paid/free per source group
source_group_perc <- data.frame(cbind(v2,v3))


visual_source_group <- melt(source_group_perc[,c('source_group', 'free_users','paid_users')],id.vars = 

p4 <- ggplot() + geom_bar(aes(y = value, x = source_group, fill=variable ), data = visual_source_group,
                          stat = "identity", position = "dodge") +
  theme(panel.background = element_rect(fill = "#31324F",
                                colour = "#31324F",
                                size = 0.5, linetype = "solid"),  axis.text.x = element_text(face="bold
        axis.text.y = element_text(face="bold", color="black", size=6)) +ylab("Fraction of free and pai

# Displaying the visualization of paid vs free users w.r.t source group
p4
```



2. Visualization of the page views per source group

Identifying the page views and the source groups can be another important link which can show exactly what pages of the applications are being view the most by each source group. It might be helpful to get several insights such as what are the top 5 or top 10 pages for for “xyz” source_group & providing more functionalities, features, advertisements on those pages compared to the other pages where there are less visits.

ANother important application can be the navigational map for the product user interface. It is challenging to come up with the ideal set of navigations for each scenario whenever a product UI is being developed. Using the insights like page view counts would definitely help by providing the quntified results to finalize the screen interaction for product user interface.

Since there were around 47 surce groups and each group containign multiple page types, I created a function called created_visualization which takes the souce group as an input and filters the dataframe with it to created the visualization of top 5 page types based on the number of views. Further, I have considered implementing the list of lists for source group and iterativelty printing the visualizations of 5 most viewd page types per souce group.

One commonly observed trend was the homepage having large number of views, which is obvious as once the application is loaded, home page would pop up first and hence, the view. What might be interesting to derive from this is the subseqquent patterns of the pages having most number of views that will lead to an indeal navigational grpah.

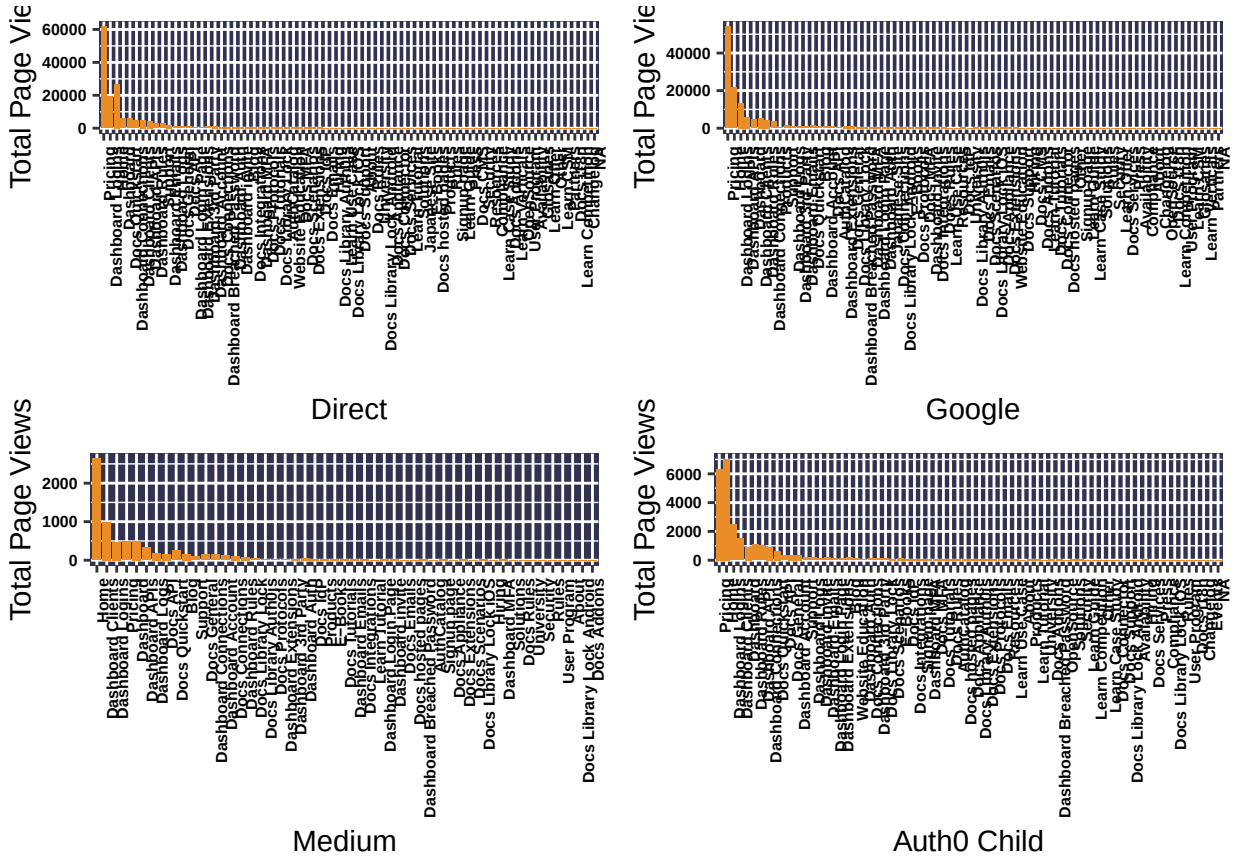
```
page_views_per_group_with_page_type
```

```
## Source: local data frame [154,250 x 14]
## Groups: source_group, type_page [1,921]
##
##    email_id invited_to_dashboard   country      role company_source
##       <chr>                <lgl>     <chr>     <chr>          <chr>
## 1         1                FALSE     Japan developer           None
## 2         1                FALSE     Japan developer           None
## 3         1                FALSE     Japan developer           None
## 4         1                FALSE     Japan developer           None
## 5         1                FALSE     Japan developer           None
## 6         1                FALSE     Japan developer           None
## 7         1                FALSE     Japan developer           None
## 8         2                 TRUE Australia               Salesforce
## 9         2                 TRUE Australia               Salesforce
## 10        2                 TRUE Australia               Salesforce
## # ... with 154,240 more rows, and 9 more variables:
## #   company_employees <int>, source_group <chr>, sperating_system <chr>,
## #   browser <chr>, dashboard_sessions <int>, login_method <chr>,
## #   is_socialmedia_influencer <lgl>, type_page <chr>, visits <dbl>
```

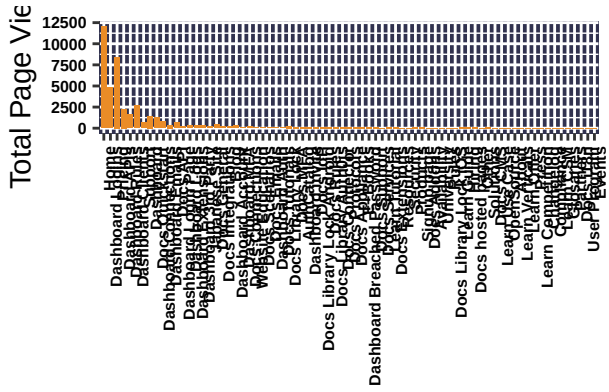
```
# function to create visualization for page views per source group
create_visualization <- function(entity){
dataframe <- page_views_per_group_with_page_type %>%
  filter(source_group == entity) %>%
  arrange(desc(visits)) %>%
  top_n(5)

 dataframe$type_page <- factor(dataframe$type_page, levels = dataframe$type_page[order(-dataframe$visits

p1 <- ggplot(dataframe, aes(x = dataframe$type_page, y = dataframe$visits)) +
  geom_bar(stat = "identity",fill = "#ea8c27") +
  theme(panel.background = element_rect(fill = "#31324F",
                                colour = "#31324F",
                                size = 0.5, linetype = "solid"),  axis.text.x = element_text(face="bold
        axis.text.y = element_text(face="bold", color="black", size=6))+ xlab(entity) +
  ylab("Total Page Views")
```

```
  return (p1)

}

# I have created nested lists where the source_type will be the input and output is list of lists where

# Getting the list of source groups
source_group_list <- as.list(unique(page_views_per_group_with_page_type$source_group))

# specifying the sublist size
k <- 4
n = length(source_group_list)

# creating the list of lists
res <- split(source_group_list, rep(1:ceiling(n/k), each=k)[1:n])

# Iterating over the list of lists and calling the grid.arrange to create the set of 4 visualization in
# iteration of the inner for loop

for (each_list in res){
  l <- list();
  i <- 1;
  for (each in each_list){

    l[[i]] <- create_visualization(each)
    i <- i + 1
  }
  do.call("grid.arrange", c(l, ncol=2))
}
```

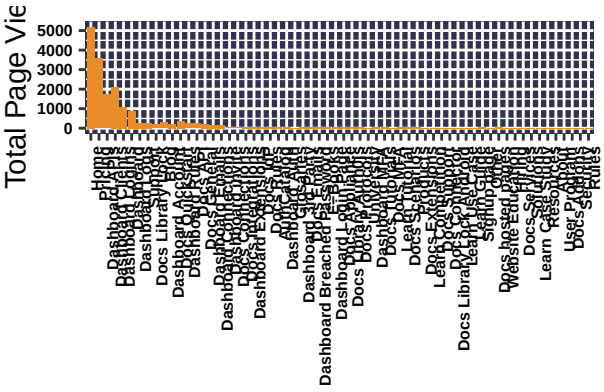
```
## Selecting by visits
## Selecting by visits
## Selecting by visits
## Selecting by visits
## Selecting by visits
## Selecting by visits
## Selecting by visits
## Selecting by visits
```



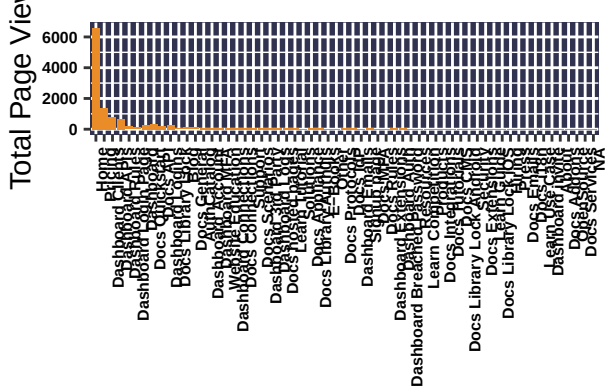

```
## Selecting by visits
## Selecting by visits
## Selecting by visits
## Selecting by visits
```



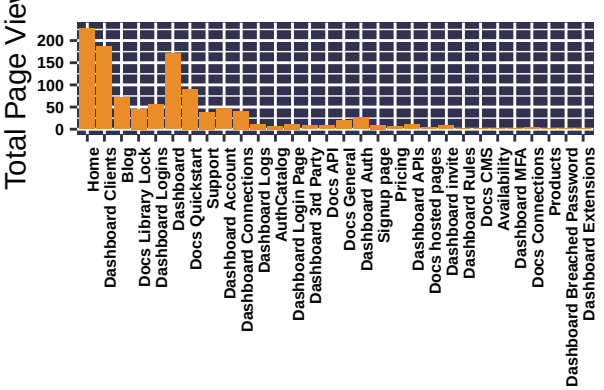
Other



Twitter

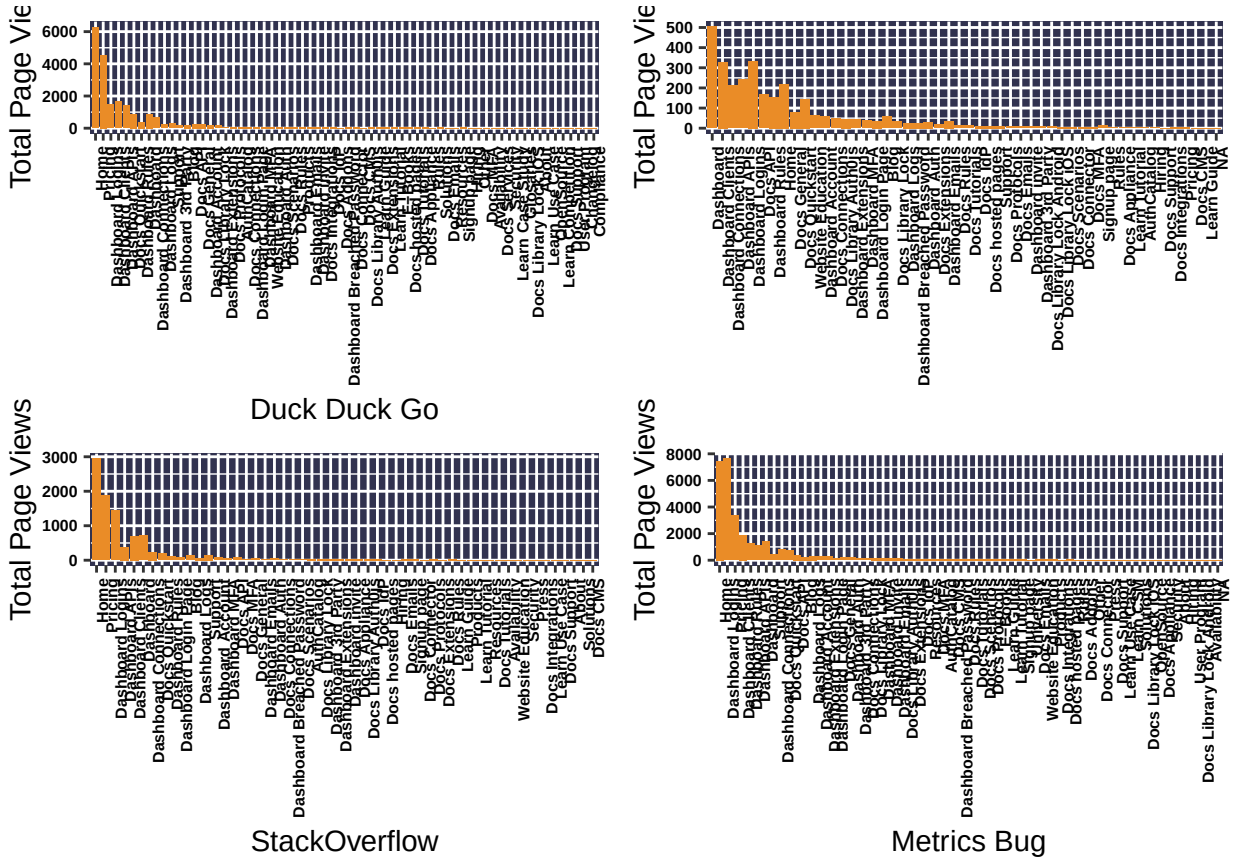


Scotch

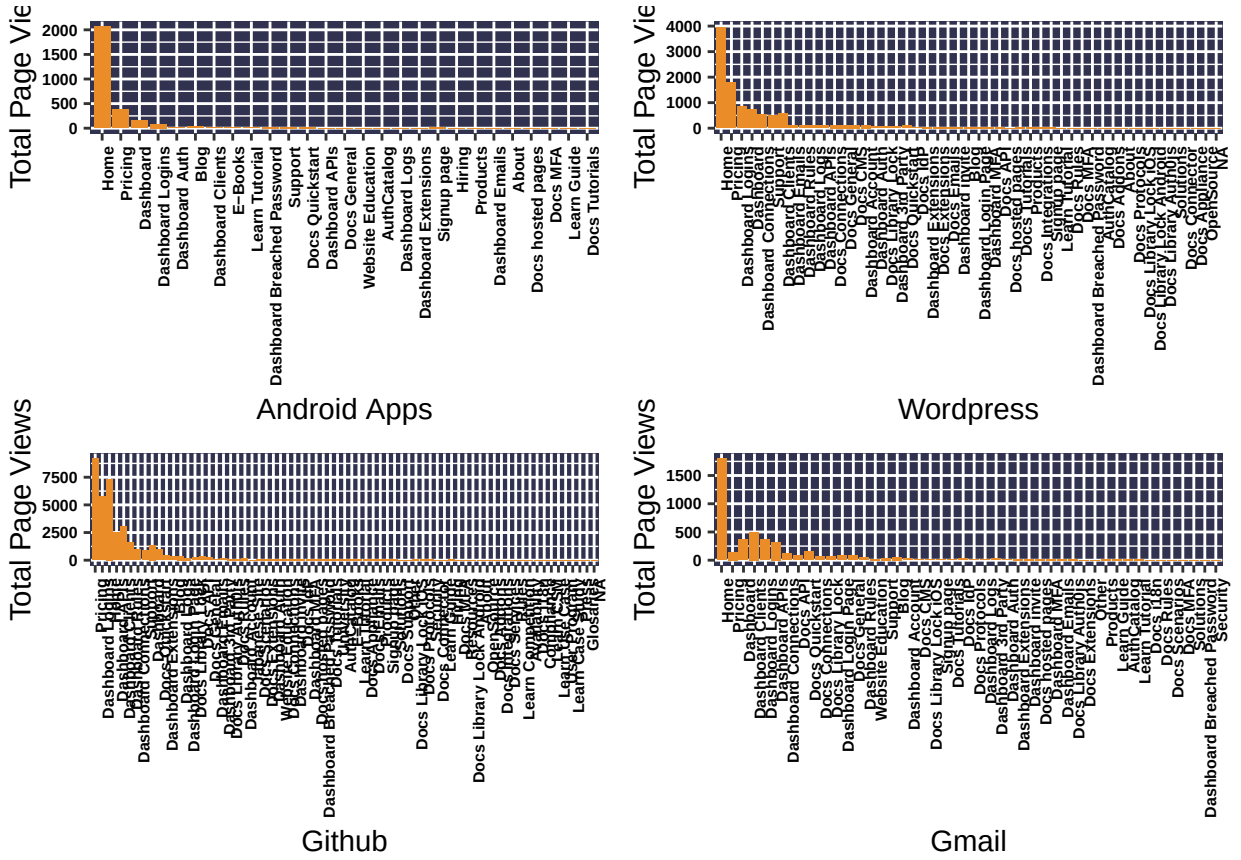


Auth0 Tenant Oops

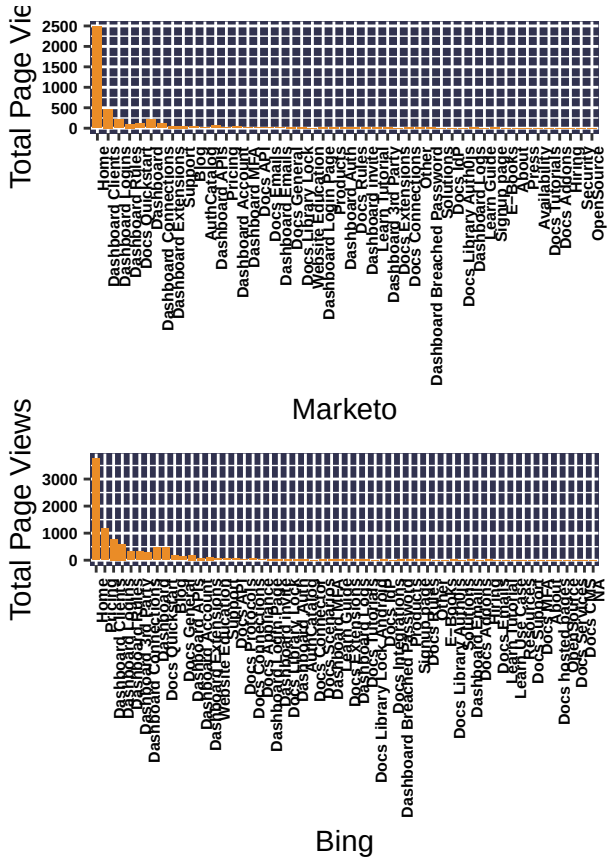
```
## Selecting by visits
## Selecting by visits
## Selecting by visits
## Selecting by visits
```

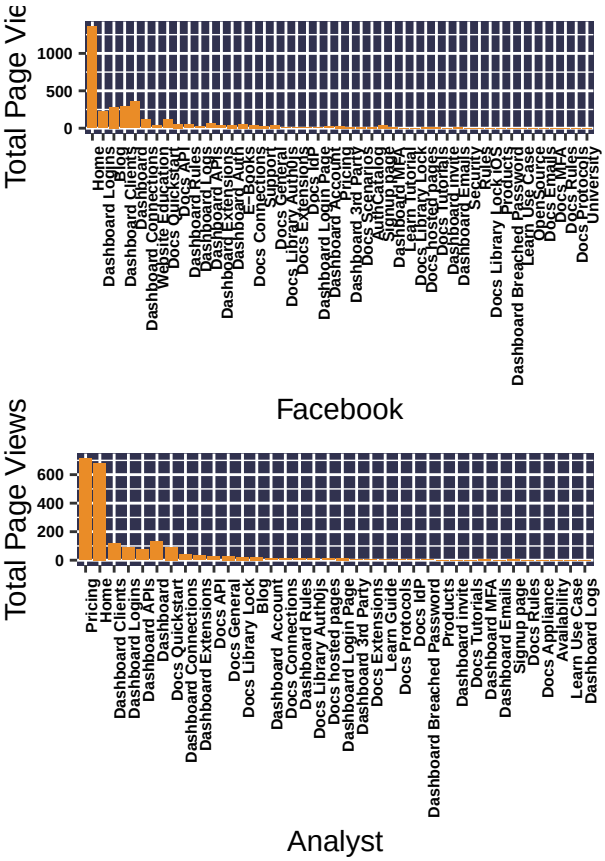



```
## Selecting by visits
## Selecting by visits
## Selecting by visits
## Selecting by visits
```

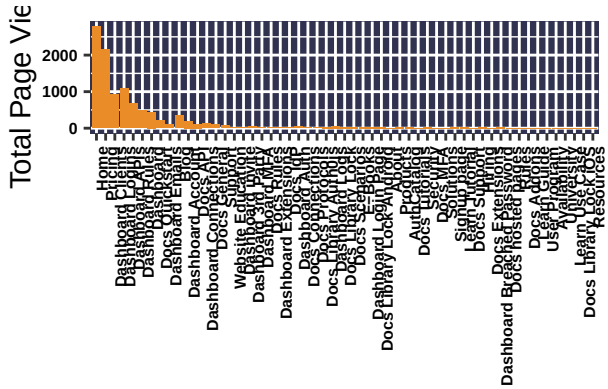



```
## Selecting by visits
## Selecting by visits
## Selecting by visits
## Selecting by visits
```

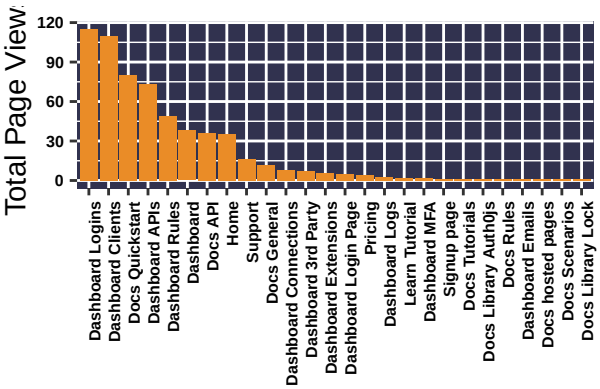




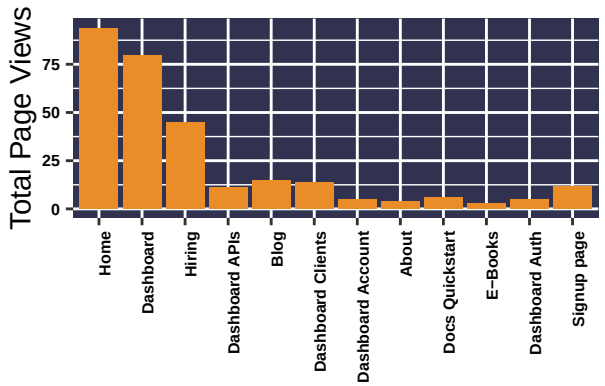

```
## Selecting by visits
## Selecting by visits
## Selecting by visits
## Selecting by visits
```



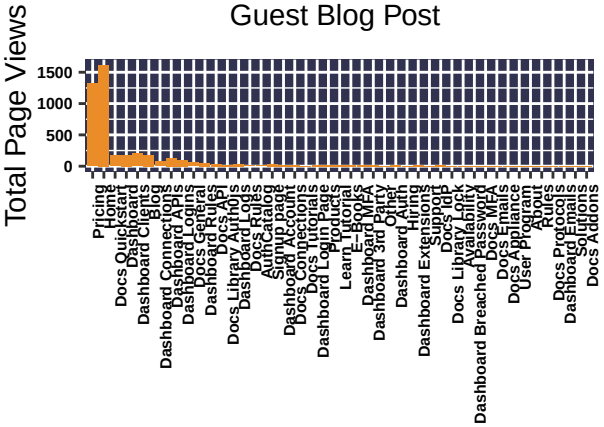
Dzone



Guest Blog Post

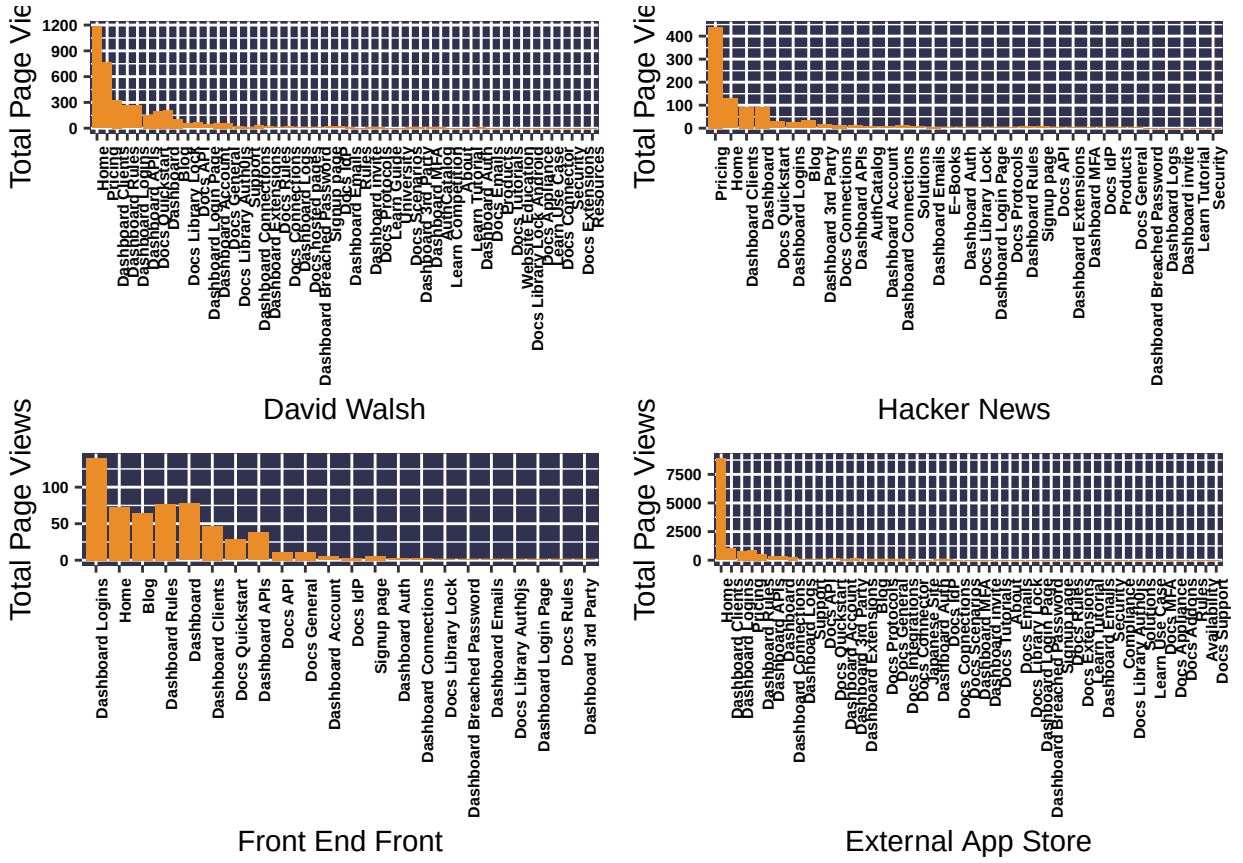


Linkedin

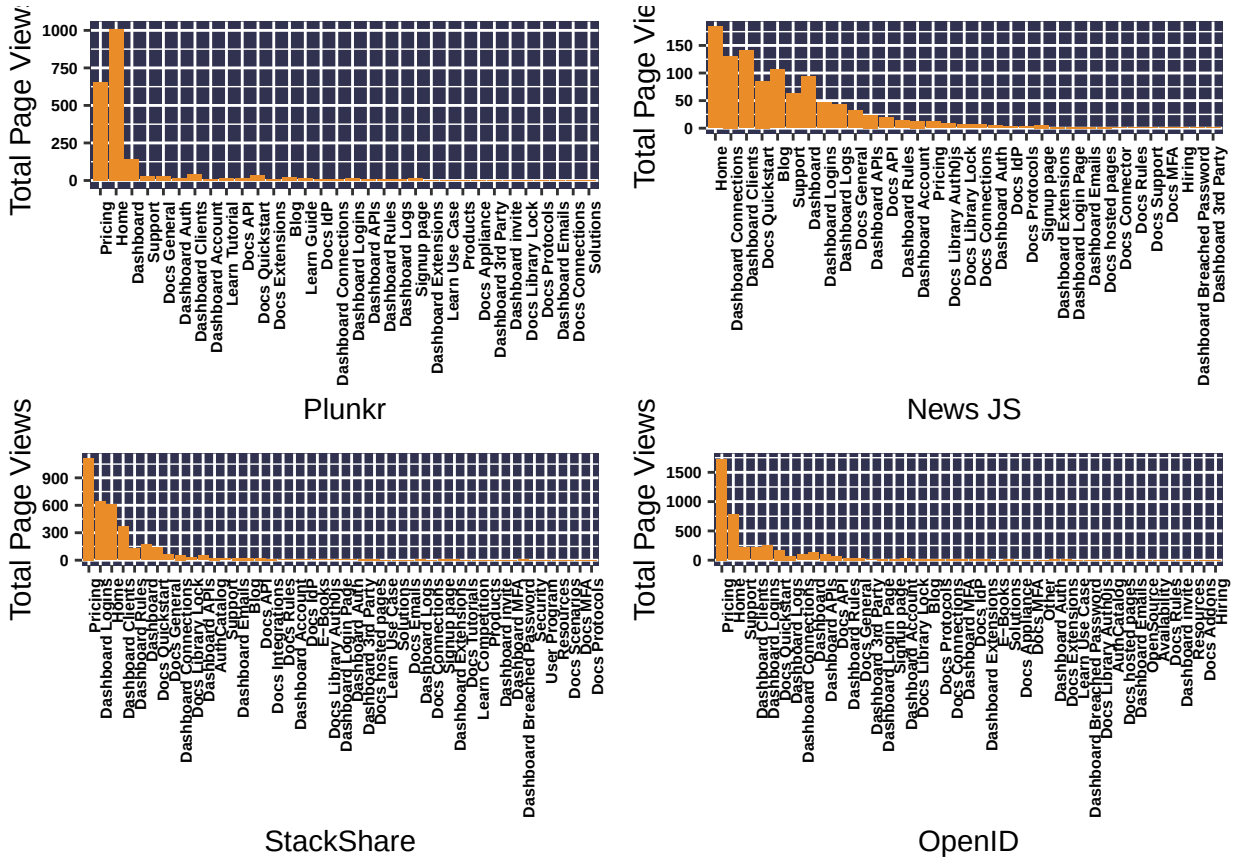


Reddit

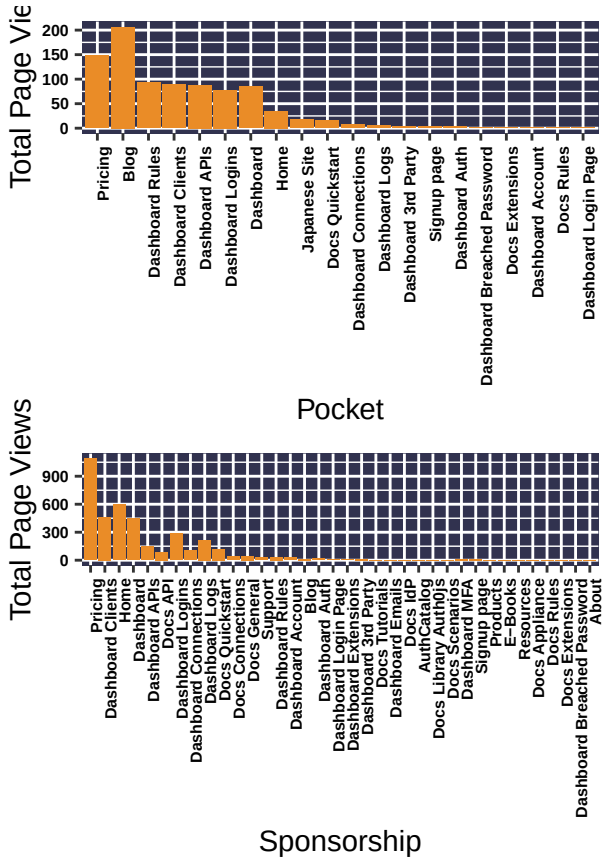
```
## Selecting by visits
## Selecting by visits
## Selecting by visits
## Selecting by visits
```

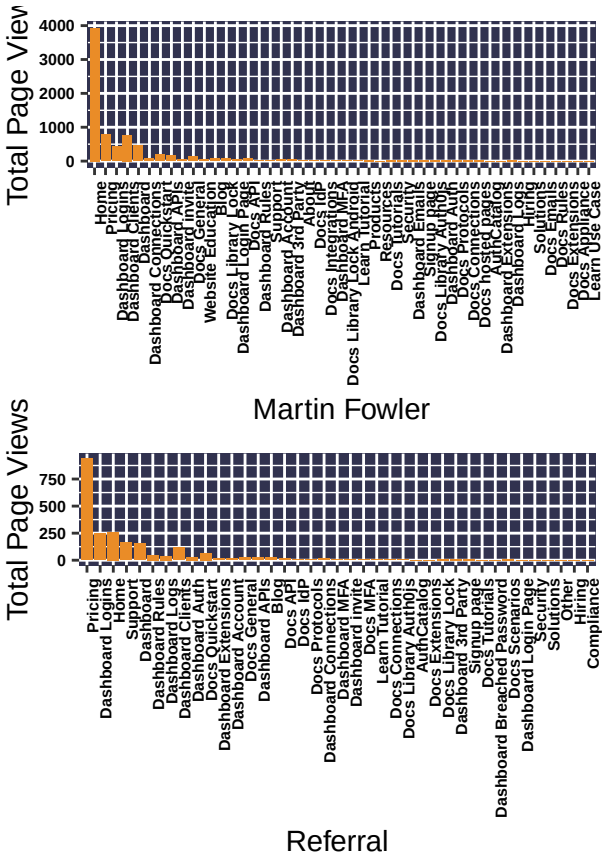



```
## Selecting by visits
## Selecting by visits
## Selecting by visits
## Selecting by visits
```

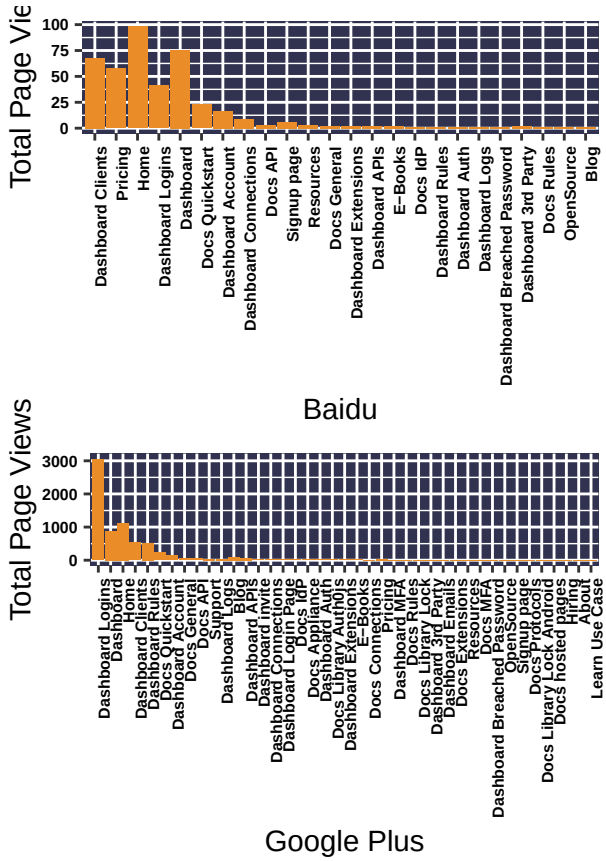


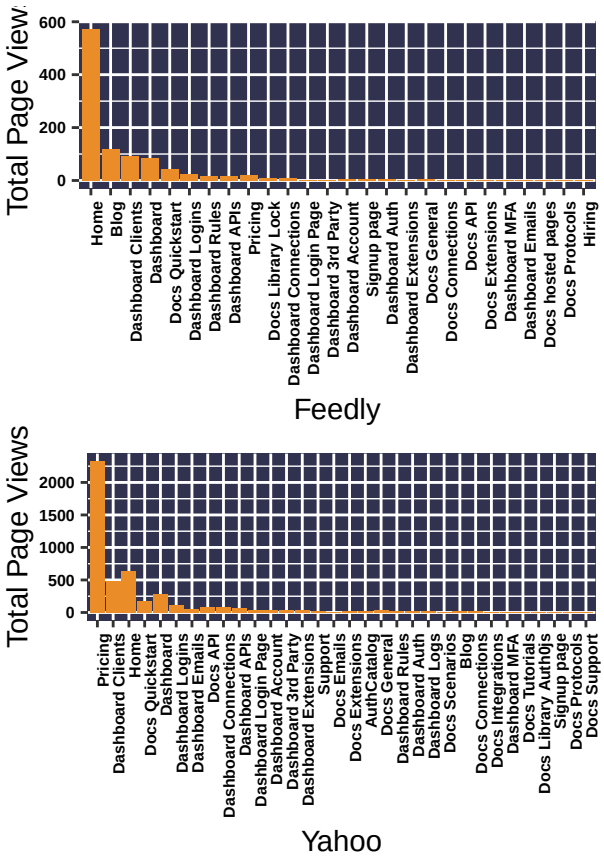

```
## Selecting by visits
## Selecting by visits
## Selecting by visits
## Selecting by visits
```



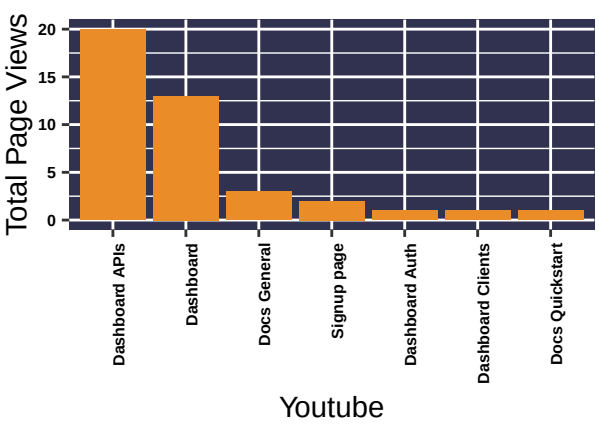



```
## Selecting by visits
## Selecting by visits
## Selecting by visits
## Selecting by visits
```

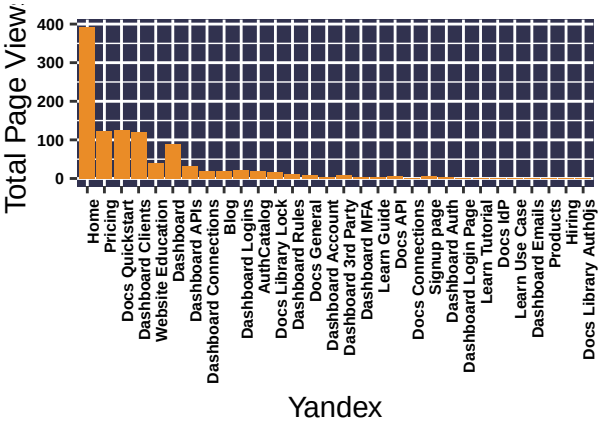




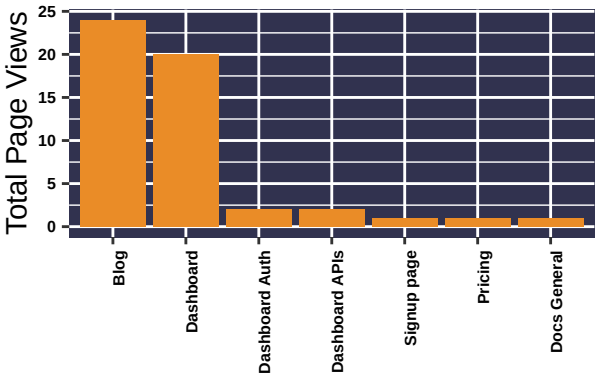

```
## Selecting by visits
## Selecting by visits
## Selecting by visits
```



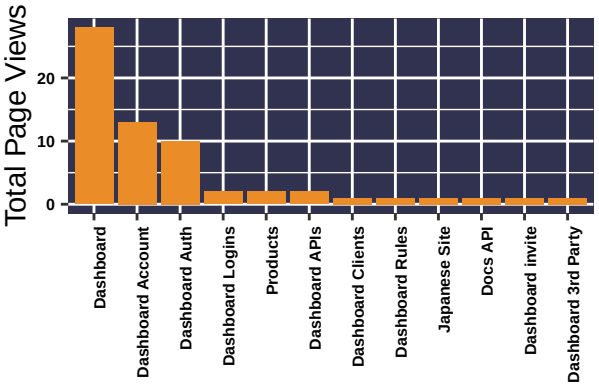
Youtube



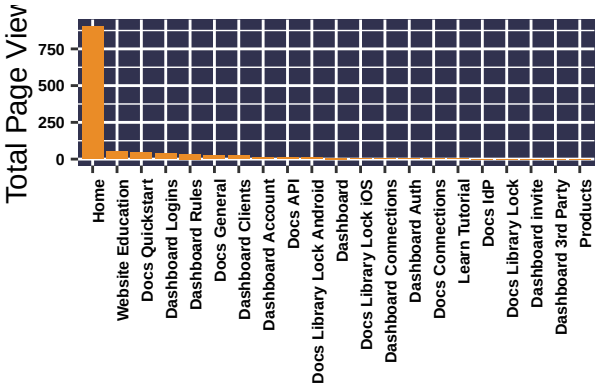
Yandex



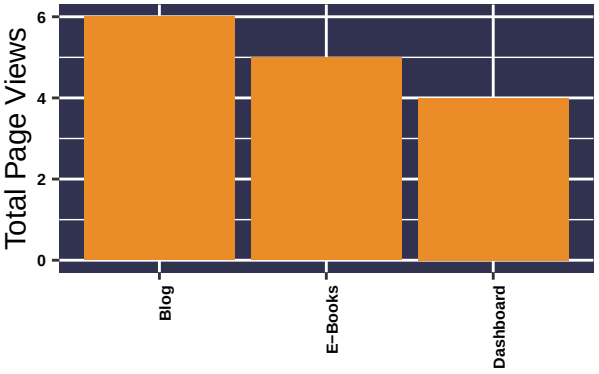
Laravel News



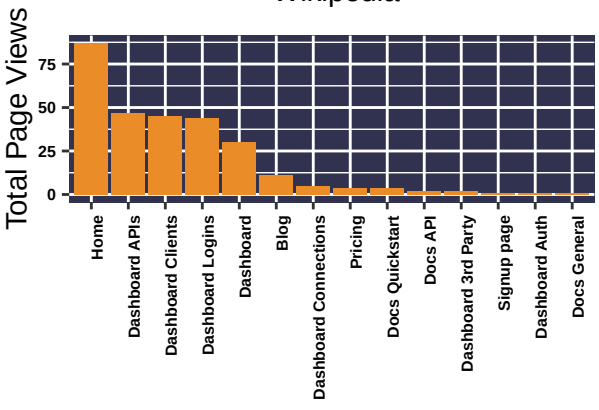
Yammer



Wikipedia



Javascript Weekly



Node Weekly

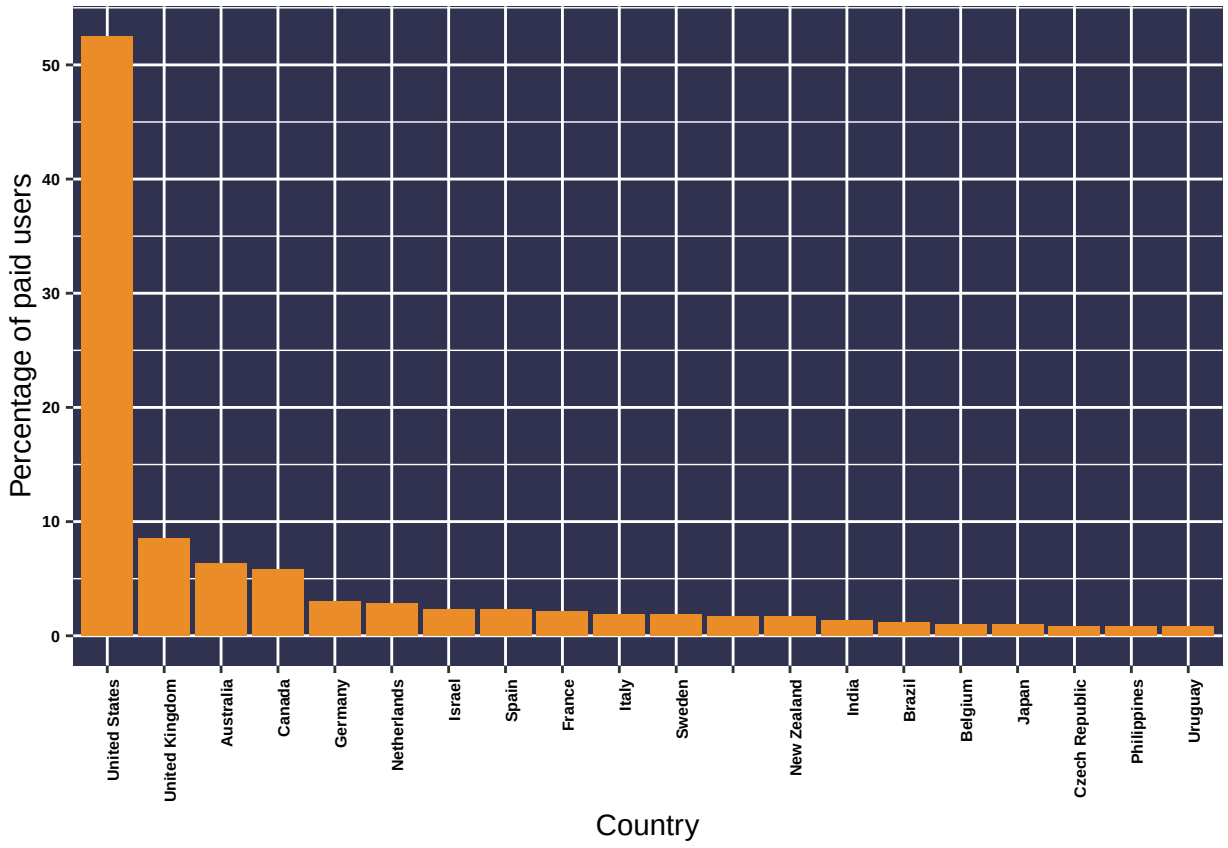
3. Visualization of the geographic information & Features

Geographic data such as region, country often provide important information such as type of user group, what are the most used features, what are the predominant technologies, etc. I have visualized some of these factors below along with the relationship between features and browser use.

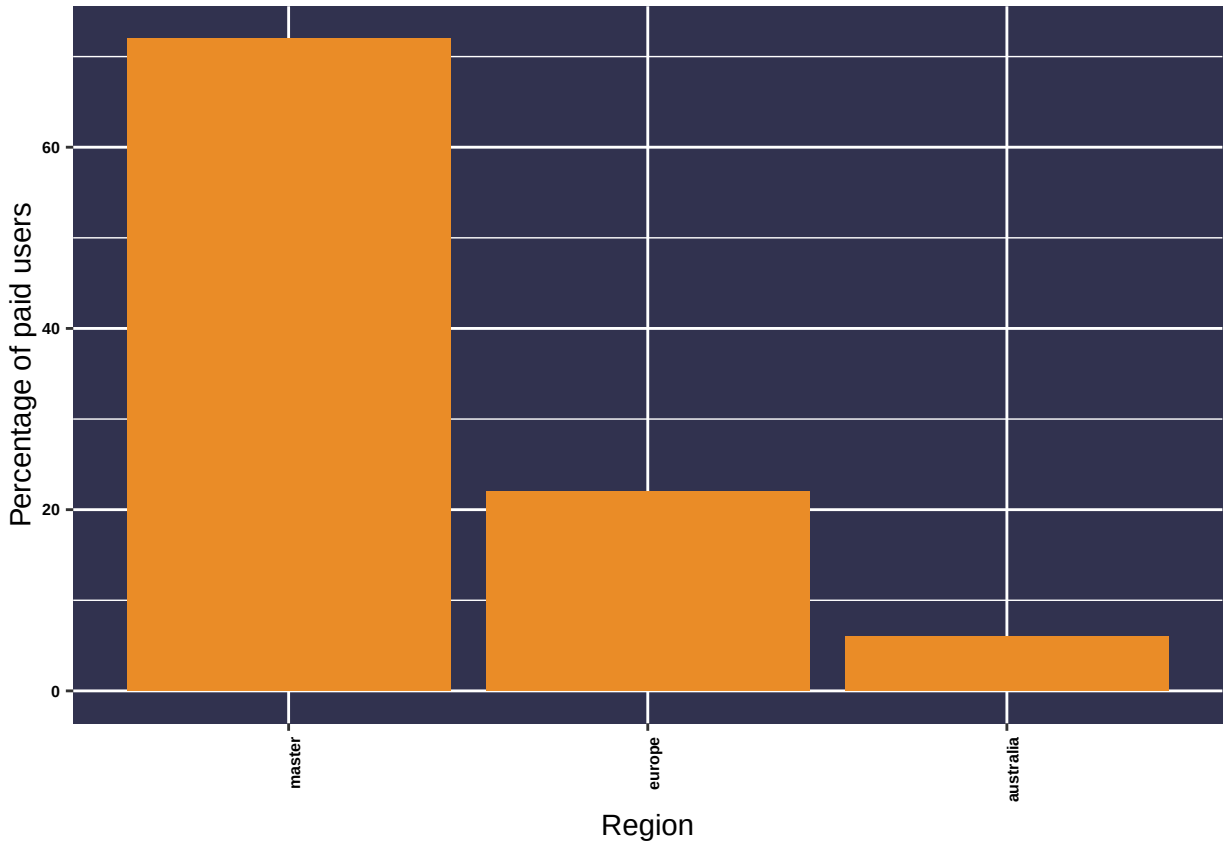
```
# 1. Countries with paid users
viz_paid_country <- dataset_tenants %>%
  left_join(dataset_relation, by="tenant_id") %>%
  left_join(dataset_emails, by="email_id") %>%
  group_by(country) %>%
  filter(paid == TRUE) %>%
  summarise(count = n()) %>%
  arrange(desc(count)) %>%
  top_n(20)
```

```
## Selecting by count
```

```
# Converting country to factors for ordered histogram and plotting the top 20 countired with highest pe
viz_paid_country$country <- factor(viz_paid_country$country, levels=viz_paid_country$country[order(-viz_
viz_paid_country$count <- viz_paid_country$count / sum(viz_paid_country$count)*100

p1 <- ggplot(viz_paid_country, aes(x = viz_paid_country$country, y = viz_paid_country$count)) +
  geom_bar(stat = "identity",fill = "#ea8c27") +
  theme(panel.background = element_rect(fill = "#31324F",
                                colour = "#31324F",
                                size = 0.5, linetype = "solid"),  axis.text.x = element_text(face="bold
        axis.text.y = element_text(face="bold", color="black", size=6))+ xlab("Country") +
  ylab("Percentage of paid users")

# Displaying the plot
p1
```




```
# 2. Regions with paid users
viz_paid_region <- dataset_tenants %>%
  left_join(dataset_relation, by="tenant_id") %>%
  left_join(dataset_emails, by="email_id") %>%
  group_by(region) %>%
  filter(paid == TRUE) %>%
  summarise(count = n()) %>%
  arrange(desc(count))

# Converting region to factors for ordered histogram and plotting results for paid users out of overall
viz_paid_region$region <- factor(viz_paid_region$region, levels=viz_paid_region$region[order(-viz_paid_
viz_paid_region$count <- viz_paid_region$count / sum(viz_paid_region$count)*100

p2 <- ggplot(viz_paid_region, aes(x = viz_paid_region$region, y = viz_paid_region$count)) +
  geom_bar(stat = "identity",fill = "#ea8c27") +
  theme(panel.background = element_rect(fill = "#31324F",
                                colour = "#31324F",
                                size = 0.5, linetype = "solid"),  axis.text.x = element_text(face="bold
        axis.text.y = element_text(face="bold", color="black", size=6))+ xlab("Region") +
  ylab("Percentage of paid users")

# Displaying the plot
p2
```




```
# 3. features w.r.t. browser

s <- strsplit(role_df_tenants_combined$used_features, split = ",")
viz_paid_features_dataset <- data.frame(tenant_id = rep(role_df_tenants_combined$tenant_id, sapply(s, le
viz_paid_features_dataset$used_features <- trim(viz_paid_features_dataset$used_features)



viz_paid_features <- dataset_emails %>%
  left_join(dataset_relation, by="email_id") %>%
  left_join(viz_paid_features_dataset, by="tenant_id")%>%
  group_by(browser, used_features) %>%
  na.omit() %>%
  summarise(count=n()) %>%
  arrange(desc(count))
```

```
## Warning in left_join_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining
## factor and character vector, coercing into character vector
```

```
viz_paid_features$count <- viz_paid_features$count / sum(viz_paid_features$count)*100

viz_paid_features$browser <- factor(viz_paid_features$browser, levels=viz_paid_features$browser[order(-v
```

```
## Warning in `levels<-`(`*tmp*`, value = if (nl == nL) as.character(labels)
## else paste0(labels, : duplicated levels in factors are deprecated
```

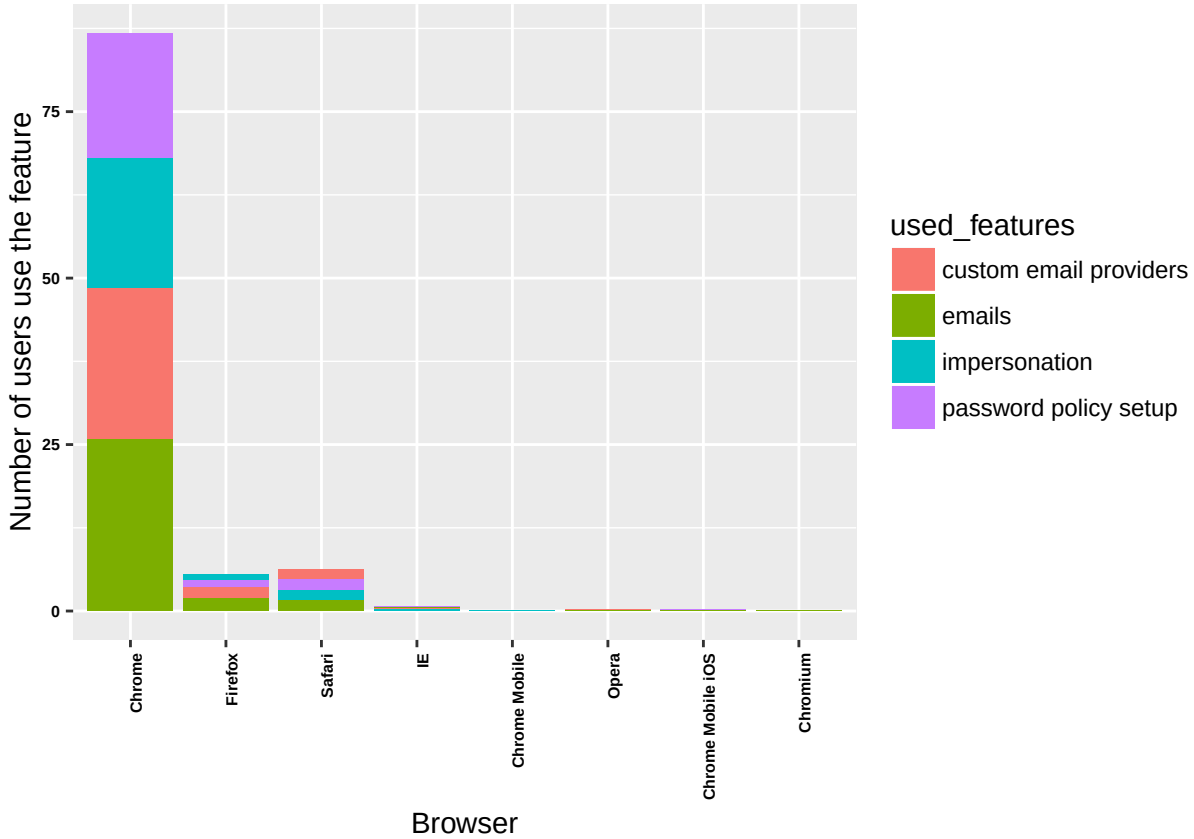
```
  geom_bar(stat = "identity")
```

```
## geom_bar: width = NULL, na.rm = FALSE
## stat_identity: na.rm = FALSE
## position_stack
```

```
p3 <- ggplot(viz_paid_features, aes(x = browser, y = count, fill = used_features)) +  geom_bar(stat = "
  theme(axis.text.x = element_text(face="bold", color="black", size=6, angle=90, hjust = 1),
        axis.text.y = element_text(face="bold", color="black", size=6)) +ylab("Number of users use the

# Displaying the visualization of broswer vs
p3
```

```
## Warning in `levels<-`(`*tmp*`, value = if (nl == nL) as.character(labels)
## else paste0(labels, : duplicated levels in factors are deprecated
```

```
## Warning in `levels<-`(`*tmp*`, value = if (nl == nL) as.character(labels)
## else paste0(labels, : duplicated levels in factors are deprecated
```



Some of the insights regaridng the paid users w.r.t geographics are: 1. United States has the highest percentage of paid users followed by United Kingdom. 2. The master region is the one with highest number of paid users 3. In terms of the features w.r.t. browser data, google chrome clearly wins as the number of features recorded for the chrome browser are highest.

Future Scope:

1. The dataset had numerious interesting entities to be identified further. For example, it would be interesting to see how the deployment type w.r.t. public vd private cloud method distributes w.r.t. open tickets. It is desirable that the public cloud should have less number of tickets when grouped under public cloud as it is harder to configure in the private cloud environmrnt due to security breach issues and limitations.

2. Having a holistic understanding of all the technical aspects together would be a great analysis task where it would be interesting to see all the technical entities like operation systems, browser type, features, connection type, etc, are related to each other along with considerations given to the demographics and users.